Estimación

jueves, 17 de mayo de 2018 10:31

Consideremos los modelos:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{i,p-1} + \varepsilon_i \quad (i=1,\dots,n) \qquad (1)$$

o

$$Y_i = \beta_0 x_{i0} + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{i,p-1} + \varepsilon_i \quad (i=1,\dots,n)$$

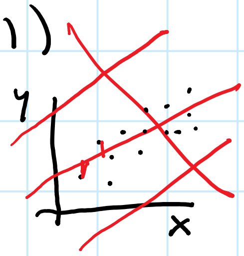
ambos modelos se pueden representar como:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} x_{10} & x_{11} & x_{12} & \cdots & x_{1,p-1} \\ x_{20} & x_{21} & x_{22} & \cdots & x_{2,p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{n0} & x_{n1} & x_{n2} & \cdots & x_{n,p-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

matricialmente:

$$Y_{n\times 1} = X_{n\times p}\, \beta_{p\times 1} + \varepsilon_{n\times 1} \qquad (2)$$

Observemos que ambos modelos en (1) se representan mediante (2) (en el primer caso $x_{i0}=1$ $(i=1,\dots,n)$)



Estimación por mínimos cuadrados

El método consiste en minimizar $\sum_{i=1}^{n} \varepsilon_i^2$ respecto a β.

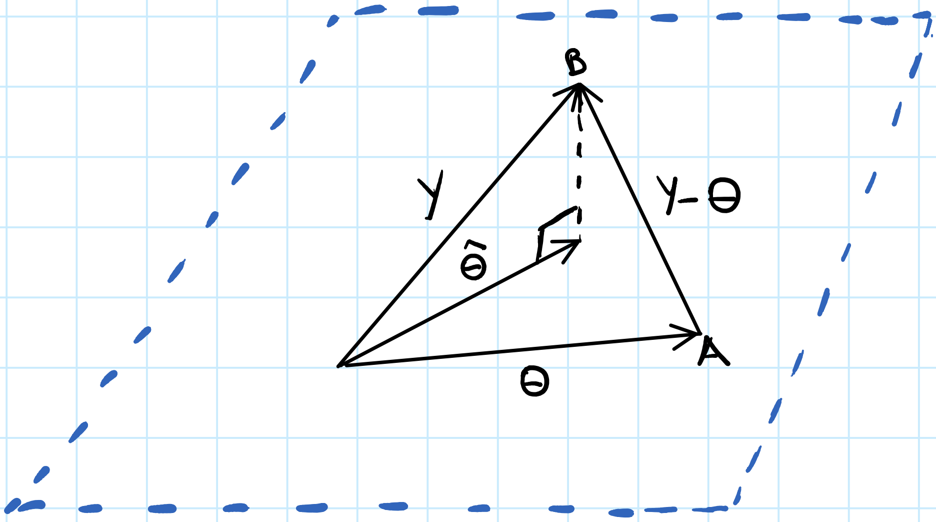
O sea se hallará $\underset{\beta}{\operatorname{argmin}}\ \varepsilon'\varepsilon = \underset{\beta}{\operatorname{argmin}} \|Y - X\beta\|^2$

Si consideramos $\Theta = X\beta$ lo anterior es equivalente a

$$\arg\min_{\theta} \|Y-\theta\|^2 \quad \text{sujeto a} \quad \theta \in \Omega$$

con $\Omega = \{y : y = Xx \text{ para todo } x \in \mathbb{R}^p\}$

(Ω es el espacio de las columnas de X)



Si θ varía en Ω, $\|Y-\theta\|^2$ será mínimo en $\theta = \hat{\theta}$ donde $\hat{\theta}$ queda determinado por $(Y-\hat{\theta}) \perp \Omega$

$\hat{\theta}$ cumple que $\hat{\theta} = PY$ donde P es la matriz de proyección ortogonal en Ω (P simétrica e idempotente)

Entonces $\quad Y-\theta = (Y-\hat{\theta}) + (\hat{\theta}-\theta)$

y como $\quad P\theta = \theta \quad$ (ya que $\theta \in \Omega$)

$P' = P \quad$ (P es simétrica)

$P^2 = P \quad$ (P es idempotente)

se tiene que:

$$(Y-\hat{\theta})'(\hat{\theta}-\theta) = (Y-PY)'(PY-P\theta) =$$

$$= (Y' - Y'\underbrace{P'}_{P})P(Y-\theta) = Y'(I_n - P)P(Y-\theta) = 0$$

La última igualdad se debe a que $(Y-\hat{\theta}) \perp (\hat{\theta}-\theta)$
(Recordar que si $a \perp b \Rightarrow a'b = 0$)

De donde $\|Y-\theta\|^2 = (Y-\theta)'(Y-\theta) =$

$$= \left((Y-\hat{\theta}) + (\hat{\theta}-\theta)\right)'\left((Y-\hat{\theta}) + (\hat{\theta}-\theta)\right) =$$

$$(Y-\hat{\theta})'(Y-\hat{\theta}) + (Y-\hat{\theta})'(\hat{\theta}-\theta) + (\hat{\theta}-\theta)'(Y-\hat{\theta}) + (\hat{\theta}-\theta)'(\hat{\theta}-\theta)$$

$$= \|Y-\hat{\theta}\|^2 + 2\underbrace{(Y-\hat{\theta})'(\hat{\theta}-\theta)}_{0} + \|\hat{\theta}-\theta\|^2$$

o sea $\|Y-\theta\|^2 = \|Y-\hat{\theta}\|^2 + \|\hat{\theta}-\theta\|^2 \geqslant \|Y-\hat{\theta}\|^2$

dándose la igualdad si y solo si $\theta = \hat{\theta}$ o sea
el mínimo se dará en $\hat{\theta}$

$Xv \in \Omega \quad \forall v \in \mathbb{R}^r$

Dado que $Y-\hat{\theta} \perp \Omega$ entonces como $Y-\hat{\theta} \in \mathrm{Ker}(X')$

$(Xv)'(Y-\hat{\theta}) = 0$

$\underbrace{v'}_{\forall v}\underbrace{X'(Y-\hat{\theta})}_{0} = 0$

$$X'(Y-\hat{\theta}) = 0$$

o

$$X'\hat{\theta} = X'Y \qquad (3)$$

De aquí $\hat{\theta}$ está determinado únicamente siendo la única proyección ortogonal de Y en Ω

Asumamos que las columnas de X son linealmente independientes así pues existe un único vector $\hat{\beta}$ tal que $\hat{\theta} = X\hat{\beta}$. Entonces a partir de

$$X'\hat{\theta} = X'Y$$

$$X'X\hat{\beta} = X'Y \qquad (4)$$

La ecuación (4) se le llama ecuaciones normales

Como X tiene rango p entonces $X'X$ es definida positiva y no singular

Así existe $(X'X)^{-1}$ y la ecuación (4) tiene solución única dada por

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

$\hat{\beta}$ se llamará el estimador de mínimos cuadrados de β

Observación: Si las columnas de X no son linealmente independientes no existe un único $\hat{\beta}$ tal que $\hat{\theta} = X\hat{\beta}$ así pues (4) no tiene una solución única. Aún así una solución esta dada por

$$\hat{\beta} = (X'X)^{-}X'Y \qquad (5)$$

siendo $(X'X)^-$ alguna inversa generalizada de $(X'X)$

Entonces $\hat{\theta} = X\hat{\beta} = X(X'X)^- X'Y = PY$
y dado que P es único, se sigue que P no depende de que inversa generalizada se usa.

Los valores $\hat{Y} = X\hat{\beta}$ se llamarán valores ajustados. (6)

A los valores $e = Y - \hat{Y}$ se les llamará residuos

$$e = (Y - \hat{Y}) = (Y - X\hat{\beta}) = (I_n - P)Y \qquad (7)$$

Vemos que el mínimo valor de $\varepsilon'\varepsilon$ se da en $e'e$

$$e'e = (Y - X\hat{\beta})'(Y - X\hat{\beta}) = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

$$= Y'Y - \hat{\beta}'X'Y + \hat{\beta}'[\underbrace{X'X\hat{\beta} - X'Y}_{0 \text{ por } (4)}] = Y'Y - \hat{\beta}'X'Y \quad (8)$$

de donde $e'e = Y'Y - \hat{\beta}X'Y = Y'Y - \hat{\beta}'X'X\hat{\beta} \quad (9)$

$e'e$ se llamará suma de cuadrados de los residuos (SCR)

Como $\hat{\theta} = X\hat{\beta}$ es único notemos que $\hat{Y}$, e y SCR son únicos independientemente del rango de X.

Teorema: Supongamos que X es una matriz $n \times p$ de rango p, y sea $P = X(X'X)^{-1}X'$
Entonces:

i) P y $I_n - P$ son simétricas e idempotentes

ii) rango $(I_n - P) = tr(I_n - P) = n - p$

iii) $PX = X$

Dem: i) $P' = (X(X'X)^{-1}X')' = X''\left[(X'X)^{-1}\right]'X' =$

$$= X\left((X'X)'\right)^{-1}X' = X(X'X)^{-1}X' = P$$

luego P es simétrica

$(I_n - P)' = I_n' - P' = I_n - P$ o sea $I_n - P$ es simétrica

$$P^2 = X\underbrace{(X'X)^{-1}X'X}_{I_p}(X'X)^{-1}X' = X I_p (X'X)^{-1}X' = P$$

$$(I_n - P)^2 = I_n - 2P + P^2 = I_n - P$$

ii) Dado que $I_n - P$ es simétrica e idempotente (prerequisitos) tenemos que

$$\text{rango}(I_n - P) = tr(I_n - P) = n - tr(P)$$

$$tr(P) = tr(X(X'X)^{-1}X') = tr(X'X(X'X)^{-1}] \circledast$$
$$= tr(I_p) = p$$

$\circledast$ prerequisitos $(tr(AC) = tr(CA))$

iii) $PX = X(X'X)^{-1}X'X = X$

<u>Teorema</u>: Sea el modelo lineal $Y = X\beta + \varepsilon$
con rango$(X) = p$

i) Si suponemos que $E(\varepsilon) = 0$ (o sea los errores son insesgados) entonces
$\hat{\beta} = (X'X)^{-1}X'Y$ es un estimador insesgado de β

ii) Si además asumimos que los errores son no correlacionados y que tienen la misma varianza, es decir $Var(\varepsilon) = \sigma^2 I_n$
entonces
$$Var(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

Dem: i) $E(\hat{\beta}) = (X'X)^{-1}X'E(Y) =$

pues $E(\varepsilon) = 0$

$= (X'X)^{-1}X'E(X\beta) = (X'X)^{-1}X'X\beta = \beta$

ii) $Var(Y) = Var(Y - X\beta) = Var(\varepsilon) = \sigma^2 I_n$

$Var(\hat{\beta}) = Var\left((X'X)^{-1}X'Y\right) =$ $\qquad V(A\cancel{Z}) = AV(Z)A'$

$= (X'X)^{-1}X'\,Var(Y)\left((X'X)^{-1}X'\right)' =$

$$= (X'X)^{-1} X' \underbrace{Var(Y)}_{\sigma^2 I_n} X (X'X)^{-1}$$

$$= \sigma^2 (X'X)^{-1} X'X (X'X)^{-1} = \sigma^2 (X'X)^{-1}$$

Consideremos $\theta = X\beta$ y sea la combinación lineal $c'\theta$. El teorema que sigue responde cual es el "mejor" (en algún sentido) estimador de esta magnitud

Teorema: Sea $\hat{\theta}$ el estimador de mínimos cuadrados de $\theta = X\beta$ con $\theta \in \Omega$ (el espacio generado por las columnas de X) y X puede no tener rango completo.
Entonces dentro de los estimadores lineales insesgados de $c'\theta$ se tiene que $c'\hat{\theta}$ es el único estimador de mínima varianza de $c'\theta$

Dem: Ya vimos que $\hat{\theta} = PY$ y

$$P\theta = PX\beta = X\beta = \theta, \text{ pues } PX = X$$

luego $E(c'\hat{\theta}) = E(c'PY) = c'P\,E(Y) =$

$$= c'P\,E(X\beta + \varepsilon) = c'PX\beta = c'P\theta = c'\theta$$

para todo $\theta \in \Omega$

o sea $c'\hat{\theta}$ es un estimador insesgado de $c'\theta$

Notemos que $c'\hat{\theta} = c'PY = (Pc)'Y$

Consideremos $d'Y$ otro estimador lineal insesgado de $c'\theta$.

Entonces: $c'\theta = E(d'Y) = d'\theta$ o sea

$$(c-d)'\theta = 0$$

luego $(c-d) \perp \Omega$

Por lo tanto $P(c-d) = 0$ de donde $Pc = Pd$

$$Var(c'\hat{\theta}) = Var((Pc)'Y) = Var((Pd)'Y) =$$

$$= (Pd)' \underbrace{Var(Y)}_{\sigma^2 I_n} (Pd) = \sigma^2 d'P'Pd = \sigma^2 d'P^2 d = \sigma^2 d'Pd$$

$d' \overset{\sigma^2 I}{Var(Y)} d = \sigma^2 d'd$

así que $Var(d'Y) - Var(c'\hat{\theta}) = Var(d'Y) - \sigma^2 d'Pd$

$$= \sigma^2 (d'd - d'Pd) = \sigma^2 d'(I_n - P)d =$$

$$= \sigma^2 d'(I_n - P)'(I_n - P)d = \sigma^2 \|(I_n - P)d\|^2 \geq 0$$

donde la igualdad se da cuando $(I_n - P)d = 0$ o sea $d = Pd = Pc$ de donde $c'\hat{\theta}$ es el estimador lineal de mínima varianza y es único.

obs.

$d'Y = (Pc)'Y =$

$= c'\underbrace{PY}_{\tilde{\theta}} = c'\hat{\theta}$

Corolario: Si X tiene rango completo, entonces $a'\hat{\beta}$ es el estimador lineal de mínima varianza de $a'\beta$ para todo vector a

Este corolario es el conocido teorema de Gauss-Markov

Estimación Insesgada de σ^2

<u>Teorema</u>: Si $E(Y)=X\beta$ donde X es una matriz $n\times p$ de rango p y $Var(Y)=\sigma^2 I_n$ entonces

$$S^2=\frac{(Y-\hat{\theta})'(Y-\hat{\theta})}{n-p}=\frac{SCR}{n-p}$$

es un estimador insesgado de σ^2.

$$Y_i=\beta_0+\varepsilon_i$$

$$\begin{pmatrix} y_1\\ \vdots \\ y_n\end{pmatrix}=\beta_0\underbrace{\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}}_{X}+\begin{pmatrix}\varepsilon_1\\ \vdots\\ \varepsilon_n\end{pmatrix}$$

$$(X'X)=n$$

$$(X'X)^{-1}=\frac{1}{n}$$

$$X'Y=\Sigma Y_i$$

$$\beta_0=\frac{\Sigma y_i}{n}$$

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

Dem: Consideremos la representación

$$\theta=X\beta$$

Entonces

$$Y-\hat{\theta}=(I_n-P)Y$$

donde $P=X(X'X)^{-1}X'$

Se tiene que

$$(n-p)S^2=Y'(I_n-P)'(I_n-P)Y=Y'(I_n-P)^2Y$$
$$=Y'(I_n-P)Y$$

En lo que sigue usaremos que si U es un vector $n\times 1$ y A una matriz simétrica $n\times n$. Si $E(U)=\mu$ y $Var(U)=\Sigma$ entonces:

$$E(U'AU) = tr(A\Sigma) + \mu' A \mu$$

Como $(I_n - P)$ es simétrica y $Var(Y) = \sigma^2 I$ y además $E(Y) = X\beta = \theta$

$$P\theta = X(X'X)^{-1}X'(X\beta) = X\beta = \theta$$

$$E(Y'(I_n-P)Y) = tr((I_n-P)\sigma^2 I_n) + E(Y)'(I_n-P)E(Y)$$

$$= \sigma^2 \underbrace{tr((I_n-P))}_{n-p} + \underbrace{\theta'(I_n-P)\theta}_{0} = \sigma^2(n-p)$$

de donde $E(S^2) = \sigma^2$

a cargo del lector

Hasta ahora las únicas hipótesis que hemos impuesto fue que $E(\varepsilon)=0$ y $Var(\varepsilon)=\sigma^2 I_n$

Si suponemos además que $\varepsilon \sim N_n(0,\sigma^2 I_n)$ obtendremos la teoría asintótica

Comencemos conociendo la distribución normal multivariada.

Definición: Sea Σ una matriz definida positiva $n\times n$ y μ un vector $n\times 1$

Llamaremos función de distribución normal multivariada aquella distribución que tenga densidad

$$f(y_1,\dots,y_n)=\frac{1}{(2\pi)^{n/2}\det(\Sigma)^{1/2}}e^{-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)}$$

Como Σ es definida positiva y por lo tanto lo es Σ^{-1} y por lo tanto la forma cuadrática $(y-\mu)'\Sigma^{-1}(y-\mu)\geq 0$

Luego $f(y_1,\dots,y_n)$ es acotada y toma su máximo en $y=\mu$ y vale $(2\pi)^{-n/2}\det(\Sigma)^{-1/2}$

Debemos ver que efectivamente f es una densidad

Como Σ es definida positiva tiene una raíz cuadrada simétrica definida positiva $\Sigma^{1/2}$ que satisface

$(\Sigma^{1/2})^2 = \Sigma$

Sea $z = \Sigma^{-1/2}(y-\mu)$ por lo que $y = \Sigma^{1/2} z + \mu$

El jacobiano de esta transformación es:

$$J = \det\left(\frac{\partial y_i}{\partial z_j}\right) = \det(\Sigma^{1/2}) = [\det(\Sigma)]^{1/2}$$

Luego

$$\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} e^{-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)}\, dy_1\ldots dy_n =$$

$$= \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} e^{-\frac{1}{2}z'\Sigma^{1/2}\Sigma^{-1}\Sigma^{1/2}z}\, |J|\, dz_1\ldots dz_n =$$

$$= \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} e^{-\frac{1}{2}z'z}\, |J|\, dz_1\ldots dz_n =$$

$$= |J| \prod_{i=1}^{n}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}z_i^2}\, dz_i = |J| \prod_{i=1}^{n}(2\pi)^{1/2} =$$

$$= (2\pi)^{n/2} \det(\Sigma)^{1/2}$$

Como $f > 0$ se sigue que es una densidad pues

$$\int_{\mathbb{R}^n} f(y_1,\ldots,y_n)\, dy_1\ldots dy_n = 1$$

<u>Teorema</u>: Si un vector aleatorio Y tiene distribución normal multivariada entonces $E(Y) = \mu$ y $Var(Y) = \Sigma$

$E(Y) = \mu$ y $Var(Y) = \Sigma$

<u>Dem</u>: Sea $Z = \Sigma^{-1/2}(Y-\mu)$

Igual que antes usando la fórmula de cambio de variable tenemos que Z tiene densidad

$$g(z_1, \dots, z_n) = f(Y(z))\,|J| = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_i^2}$$

$$= \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}z'z}$$

Por lo tanto la factorización anterior implica que los Z_i son independientes con $Z_i \sim N(0,1)$.

Esto es $E(Z) = 0$ y $Var(Z) = I_n$ luego

$$E(Y) = E(\Sigma^{1/2}Z + \mu) = \Sigma^{1/2}\overset{0}{\overset{\shortparallel}{E(Z)}} + \mu = \mu$$

y $$Var(Y) = Var(\Sigma^{1/2}Z + \mu) = Var(\Sigma^{1/2}Z) =$$

$$= \Sigma^{1/2}\,Var(Z)\,\Sigma^{1/2} = \Sigma^{1/2} I_n \Sigma^{1/2} = \Sigma$$

Para indicar que Y tiene distribución normal multivariada con $E(Y) = \mu$ y $Var(Y) = \Sigma$ utilizaremos la notación $Y \sim N(\mu, \Sigma)$

Si $Y \sim N(\mu, \Sigma)$ vimos que $Z = \Sigma^{-1/2}(Y-\mu)$ tiene $Z \sim N(0, I_n)$. Z es el análogo de la estandarización en una dimensión

Función Generatriz de Momentos

Definición: Si X es un vector aleatorio $n\times 1$ y t un vector de constantes $n\times 1$ definimos la función generatriz de momentos de X mediante

$$M_X(t) = E(e^{t'X})$$

Un resultado importante es que si $M_X(t)$ existe para todo t, tal que $\|t\| \leq t_0$ $(t_0 > 0)$ entonces $M_X(t)$ determina la distribución de X

Sea $X_1 = (x_1, \dots, x_r)'$ y $X_2 = (x_{r+1}, \dots, x_n)'$
Se prueba que X_1 y X_2 son independientes si y sólo si $M_X(t) = a(t_1, \dots, t_r)\, b(t_{r+1}, \dots, t_n)$
para algunas funciones $a(\cdot)$ y $b(\cdot)$
También lo serán si

$$M_X(t) = M_X(t_1, \dots, t_r, 0, \dots, 0)\, M_X(0, \dots, 0, t_{r+1}, \dots, t_n)$$

Si $Z \sim N_n(0, I_n) \Rightarrow E(e^{t'Z}) = e^{\frac{1}{2}t't}$

Si $Y \sim N_n(\mu, \Sigma) \Rightarrow E(e^{t'Y}) = e^{t'\mu + \frac{1}{2}t'\Sigma t}$

Veremos la función generatriz de la normal multivariada.

Primero consideremos el caso $Z \sim N(0, I_n)$ entonces por la independencia de las Z_i la función generatriz de momentos de Z es:

$$E(e^{t'Z}) = E(e^{\sum_{i=1}^n t_i Z_i}) = E(\prod_{i=1}^n e^{t_i Z_i})$$

$$= \prod_{i=1}^n E(e^{t_i Z_i}) = \prod_{i=1}^n e^{t_i^2/2} = e^{\frac{1}{2}t't}$$

donde hemos usado que si $Z_i \sim N(0,1) \Rightarrow$

$$E(e^{t_i Z_i}) = e^{t_i^2/2}$$

Si $Y \sim N_n(\mu, \Sigma)$ podemos escribir $Y = \Sigma^{1/2} Z + \mu$

con $Z \sim N_n(0, I_n)$. Si ponemos $s = \Sigma^{1/2} t$

tenemos

$$E(e^{t'Y}) = E(e^{t'(\Sigma^{1/2} Z + \mu)}) =$$

$$= E(e^{s'Z})\, e^{t'\mu} = e^{\frac{1}{2}s's}\, e^{t'\mu} =$$

$$= e^{\frac{1}{2}t'\Sigma^{1/2}\Sigma^{1/2}t + t'\mu} = e^{t'\mu + \frac{1}{2}t'\Sigma t}$$

$$\boxed{\text{Teorema: Sea } Y \sim N_n(\mu, \Sigma) \text{ y } C \text{ una matriz } m\times n \text{ de rango } m \text{ y } d \text{ un vector } m\times 1. \text{ Entonces } CY + d \sim N_m(C\mu + d,\ C\Sigma C')}$$

Dem: La función generatriz de momentos de $CY+d$ es:

$$E(e^{t'(CY+d)}) = E(e^{(C't)'Y + t'd}) =$$

$$E(e^{t'(CY+d)}) = E(e^{(C't)'Y + t'd}) =$$

$$= e^{(C't)'\mu + \frac{1}{2}(C't)'\Sigma C't + t'd} = e^{t'(C\mu+d) + \frac{1}{2}t'C\Sigma C't}$$

Dado que $C\Sigma C'$ es definida positiva la ecuación de arriba es la función generatriz de momentos de $N_m(C\mu+d, C\Sigma C')$. Observemos que $C\Sigma C'$ es definida positiva ya que Σ lo es y C es de rango m

<u>Corolario</u> Si $Y = AZ+\mu$ con A una matriz $n\times n$ no singular entonces $Y \sim N_n(\mu, AA')$

Necesitaremos a veces trabajar con vectores aleatorios de la forma CY donde Y es normal pero C no es de rango completo, por ejemplo vectores de residuos y de ajustes.

Si $Z \sim N_m(0, I_m)$ y A una matriz $n\times m$, μ un vector $n\times 1$ reemplazando $\Sigma^{1/2}$ por A en la obtención de la matriz generadora de momentos de la normal se obtiene que la función generatriz de momentos de $Y = AZ+\mu$ es:

$$e^{t'\mu + \frac{1}{2}t'AA't}$$

Notemos que $E(Y) = AE(Z)+\mu = \mu$, $Var(Y) = A\,Var(Z)A' = AA'$

Estos resultados motivan definir:

Definición: Un vector aleatorio Y, $n\times 1$ con media μ y varianza Σ, tiene una distribución normal multivariada si tiene la misma distribución de $AZ+\mu$ donde A es cualquier matriz $n\times m$ que satisface $\Sigma = AA'$ y $Z\sim N_m(0, I_m)$. Escribiremos que $Y \sim AZ+\mu$ para indicar que Y y $AZ+\mu$ tienen la misma distribución.

Es fácil verificar que si Σ es definida positiva la nueva definición es equivalente a la vieja.

Si Σ es de rango $m<n$ la distribución de Y no puede ser expresada en términos de la función de densidad. En ambos casos la función generatriz de momentos es

$$e^{t'\mu + \frac{1}{2}t'\Sigma t}$$

Escribiremos $Y\sim N_n(\mu, \Sigma)$. Cuando Σ no tiene rango completo diremos a veces que Y tiene distribución singular.

Teorema: Un vector aleatorio Y con matriz de varianza covarianza Σ, vector de medias μ tiene distribución normal $N_n(\mu, \Sigma)$ si y sólo si $a'Y$ tiene distribución normal univariada para todo vector a

Dem: Se omite.

Independencia Estadística

Ya sabemos que si tenemos un par de variables independientes son no correlacionadas. Para el caso de la distribución normal el recíproco es verdadero.

<u>Teorema</u>: Sea $Y \sim N_n(\mu, \Sigma)$ y que Y, μ y Σ se pueden particionar como:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

de manera conformable.

Entonces Y_1 y Y_2 son independientes si y solo si $\Sigma_{12} = 0$

<u>Dem</u>: La f.g.m de Y es $e^{t'\mu + \frac{1}{2}t'\Sigma t}$

Particionamos t conformablemente con Y. Entonces

$$t_1'\mu_1 + t_2'\mu_2 + \frac{1}{2}t_1'\Sigma_{11}t_1 + \frac{1}{2}t_2'\Sigma_{22}t_2 + t_1'\Sigma_{12}t_2$$

Si $\Sigma_{12} = 0$ entonces la función generatriz de momentos puede ser factorizada en términos de t_1 y t_2 separadamente lo que implica que Y_1 y Y_2 son independientes

Inversamente si Y_1 e Y_2 son independientes $\Rightarrow$

$$M(t_1, 0)\, M(0, t_2) = M(t_1, t_2)$$

donde M es la f.g.m de Y.

Por lo anterior esto implica que $t_1'\Sigma_{12}t_2 = 0$ para todo t_1 y t_2 lo cual implica que $\Sigma_{12} = 0$ (esto se sigue poniendo $t_1 = (1, 0, \ldots, 0)'$ etc)

Teorema: Sea $Y \sim N_n(\mu, \Sigma)$ y sea $U = AY$, $V = BY$
Entonces U y V son independientes si y sólo si
$Cov(U, V) = A \Sigma B' = 0$

Dem: A cargo del lector.

Distribución de los Estimadores

Teorema: Si $Y \sim N_n(X\beta, \sigma^2 I_n)$ donde X es $n \times p$ de rango p entonces

i) $\hat{\beta} \sim N_p(\beta, \sigma^2 (X'X)^{-1})$

ii) $\dfrac{(\hat{\beta} - \beta)' X'X (\hat{\beta} - \beta)}{\sigma^2} \sim \chi^2_p$

iii) $\hat{\beta}$ es independiente de S^2

iv) $\dfrac{RSS}{\sigma^2} = \dfrac{(n-p) S^2}{\sigma^2} \sim \chi^2_{n-p}$

Dem: i) Dado que $\hat{\beta} = (X'X)^{-1} X' Y = CY$

donde C es una matriz $p \times n$ tal que

$rango(C) = rango(X') = rango\, X = p$

$(\text{rango}(\Delta) = \text{rango}(\Delta') = \text{rango}(\Delta'\Delta))$

Entonces $\hat{\beta}$ tiene distribución normal multivariada con $E(\hat{\beta}) = \beta$ y $\text{Var}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$

ii) $(\hat{\beta}-\beta)' \frac{X'X}{\sigma^2}(\hat{\beta}-\beta) = (\hat{\beta}-\beta)'\left(\text{Var}(\hat{\beta})\right)^{-1}(\hat{\beta}-\beta)$ la cual tiene distribución χ^2_p por i) y porque si

$Y \sim N_n(\mu, \Sigma)$ con Σ definida positiva $\Rightarrow$
$Q = (Y-\mu)' \Sigma^{-1} (Y-\mu) \sim \chi^2_n$ (veremos su demostración más adelante)

iii) $\text{Cov}(\hat{\beta}, Y - X\hat{\beta}) = \text{Cov}((X'X)^{-1}X'Y, (I_n - P)Y)$

$$= (X'X)^{-1} X' \text{Cov}(Y,Y) (I_n - P)' = \sigma^2 (X'X)^{-1} X' (I_n - P)$$
$$= \sigma^2 ((X'X)^{-1}X' - (X'X)^{-1}X'P) =$$
$$= \sigma^2 ((X'X)^{-1}X' - \overbrace{(X'X)^{-1}X'X}^{I}(X'X)^{-1}X') = 0$$

Esto implica que $\hat{\beta}$ y $Y - X\hat{\beta}$ son independientes por lo que $\hat{\beta}$ y $\|Y - X\hat{\beta}\|^2$ son independientes o sea $\hat{\beta}$ y S^2 son independientes.

iv) $RSS = Y'(I_n - P)Y = \underbrace{(Y - X\beta)'(I_n - P)(Y - X\beta)}_{*}$

$$= \varepsilon'(I_n - P)\varepsilon$$

b) $(I_n - P)X\beta = X\beta - \underbrace{PX\beta}_{X\beta} = 0$ y $(X\beta)'(I_n-P) = ((I_n-P)X\beta)' = 0$

c) $\beta' X' \underbrace{(I_n - P) X\beta}_{0} = 0$

Tenemos que $RSS = \varepsilon'(I_n - P)\varepsilon$ con

$I_n - P$ idempotente, simétrica y de rango $n-p$

Además $\varepsilon \sim N_n(0, \sigma^2 I_n)$

Si aplicamos el resultado que dice que si $Y \sim N_n(0, I_n)$ y A una matriz simétrica entonces $Y'AY \sim \chi^2_r$ si y solo si A es idempotente de rango r (lo demostraremos más adelante) obtenemos que $\frac{RSS}{\sigma^2} \sim \chi^2_{n-p}$

Estimación Máximo Verosimil