Distribución Normal Multivariada

Raúl Alberto Pérez
raperez1@unal.edu.co
Profesor Asociado - Escuela de Estadística
Universidad Nacional de Colombia, Sede Medellín

Semestre 2022-01

Distribución Normal Multivariada

Normal Univariada: Una v.a $X \sim N(\mu, \sigma^2)$, si su f.d.p es:

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

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2}} e^{-\frac{1}{2}(x-\mu)(\sigma^2)^{-1}(x-\mu)}$$

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

donde, $E[X] = \mu$ y $Var[X] = \sigma^2$.

NOTA: Notar que $\left(\frac{X-\mu}{\sigma}\right)^2 = (X-\mu)^t(\sigma^2)^{-1}(X-\mu)$, es el cuadrado de la distancia entre X y μ escalada por su desviación estándar.

Normal Multivariada

En el caso multivariado, se trabaja con la distancia de Mahalanobis entre el vector aleatorio $\underline{\boldsymbol{x}}$ y su vector de media poblacional $\underline{\mu}$, es decir, con:

$$(\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}})^t \boldsymbol{\Sigma}^{-1} (\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}}),$$

donde, $E[\underline{\boldsymbol{x}}] = \underline{\mu}$ y $Var[\underline{\boldsymbol{x}}] = \boldsymbol{\Sigma}$.

Definicion: f.d.p Normal

Sea $\underline{\boldsymbol{x}}$ un vector aleatorio p-dimensional, ie. $\underline{\boldsymbol{x}} \in \boldsymbol{R}^p$. Se dice que $\underline{\boldsymbol{x}}$ tiene una distribución aleatoria Normal-Multivariada (o normal p-variada), con vector de medias $\underline{\mu}$ y matriz de Var-Cov $\boldsymbol{\Sigma}$, lo cual se denota por:

$\underline{\boldsymbol{x}} \sim N_p(\underline{\mu}\ ,\ \boldsymbol{\Sigma})$, si su f.d.p multivariada está dada por:

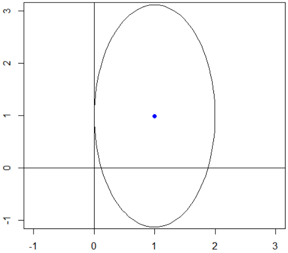
$$f(\underline{\boldsymbol{x}}) = (2\pi)^{-\frac{p}{2}} \times |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \times e^{-\frac{1}{2}(\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}})^t \boldsymbol{\Sigma}^{-1} (\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}})}.$$

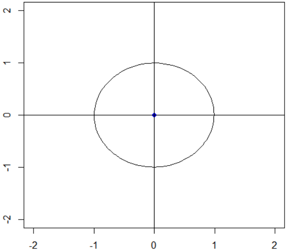
Algunos Aspectos Geométricos de la Normal Multivariada

La expresión $(\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}})^t \boldsymbol{\Sigma}^{-1} (\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}}) = c^2$, en la cual se basa el exponente de la f.d.p. normal multivariada, corresponde a un hiper-elipsoide, para culaquier contante $c > 0$, el cual está centrado en $\underline{\mu}$ y sus ejes están dados por: $\pm c\sqrt{\lambda_i}\underline{\boldsymbol{e}}_i$ para $i = 1, 2, \ldots, p$, donde.los λ_i-son los valores propios de $\boldsymbol{\Sigma}$ asociados a los respectivos vectores propios $\underline{\boldsymbol{e}}_i$.

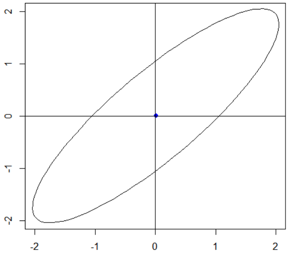
Para el caso de $p = 3$-un Elipsoide, para el caso de $p = 2$-una Elipse.

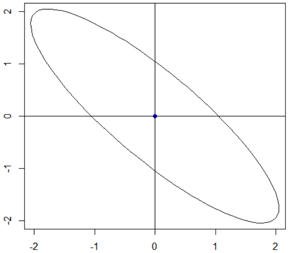
Ejemplo-1: Aspecto Gráfico





$$\underline{\mu} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{S} = \begin{bmatrix} 0.165 & 0 \\ 0 & 0.75 \end{bmatrix} \qquad \underline{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{S} = \begin{bmatrix} 0.165 & 0 \\ 0 & 0.165 \end{bmatrix}$$





$$\underline{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{S} = \begin{bmatrix} 0.7 & 0.6 \\ 0.6 & 0.7 \end{bmatrix} \qquad \underline{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{S} = \begin{bmatrix} 0.7 & -0.6 \\ -0.6 & 0.7 \end{bmatrix}$$

Ejemplo-2: Aspecto Gráfico

Suponga que un vector aleatorio 2-dimensional $\underline{\boldsymbol{x}}$ tiene la siguiente matriz de Var-Cov,

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{11}, \end{bmatrix}$$

es decir que: $\sigma_{11} = \sigma_{22}$. Graficar el elipsoide (la elipse) correspondiente bajo la restricción de que $\rho = Corr(X, Y) > 0$. Es decir, dos variables con igual varianza y correlación positiva.

Solución:

Los ejes de la elipse están dados por: $\pm c\sqrt{\lambda_i}\underline{\boldsymbol{e}_i}$, para $i = 1, 2$ y $c > 0$, es decir: $\pm c\sqrt{\lambda_1}\underline{\boldsymbol{e}_1}$ y $\pm c\sqrt{\lambda_2}\underline{\boldsymbol{e}_2}$.

Ahora se deben hallar los valores propios de $\mathbf{\Sigma}$, para lo cual se debe resolver la ecuación característica dada por:

$$|\mathbf{\Sigma} - \lambda I_2| = 0.$$

$$|\mathbf{\Sigma} - \lambda I_2| = \left| \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{11} \end{pmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right| = \begin{vmatrix} \sigma_{11} - \lambda & \sigma_{12} \\ \sigma_{12} & \sigma_{11} - \lambda \end{vmatrix}$$

$$= (\sigma_{11} - \lambda)^2 - \sigma_{12}^2$$

$$= (\sigma_{11} - \lambda - \sigma_{12})(\sigma_{11} - \lambda + \sigma_{12})$$

Al hacer, $(\sigma_{11} - \lambda - \sigma_{12})(\sigma_{11} - \lambda + \sigma_{12}) = 0$ se obtiene que:

$$\lambda_1 = \sigma_{11} + \sigma_{12} \text{ y que } \lambda_2 = \sigma_{11} - \sigma_{12}.$$

Ahora los vectores propios asociados a estos valores propios se obtienen resolviendo la ecuación:

$$\boldsymbol{\Sigma}\underline{\boldsymbol{e}} = \lambda\underline{\boldsymbol{e}}$$

Por Ejemplo si,

$\underline{\boldsymbol{e}}_1 = \begin{bmatrix} e_{11} \\ e_{12} \end{bmatrix}$-es el vector propio asociado al valor propio $\lambda_1 = \sigma_{11} + \sigma_{12}$, entonces debe cumplir que:

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{11} \end{pmatrix} \begin{pmatrix} e_{11} \\ e_{12} \end{pmatrix} = (\sigma_{11} + \sigma_{12}) \begin{pmatrix} e_{11} \\ e_{12} \end{pmatrix},$$

de donde, se obtiene el siguiente sistema de ecuaciones:

$$\sigma_{11}e_{11} + \sigma_{12}e_{12} = (\sigma_{11} + \sigma_{12})e_{11} \quad \Longrightarrow \quad e_{11} = e_{12}$$

$$\sigma_{12}e_{11} + \sigma_{11}e_{12} = (\sigma_{11} + \sigma_{12})e_{12} \quad \Longrightarrow \quad e_{11} = e_{12}$$

El vector,

$$\underline{\boldsymbol{e}}_1 = \begin{bmatrix} e_{11} \\ e_{12} \end{bmatrix} = \begin{bmatrix} e_1 \\ e_1 \end{bmatrix} \qquad \text{normalizado está dado por:}$$

$$\frac{\underline{\boldsymbol{e}}_1}{\|\underline{\boldsymbol{e}}_1\|} = \frac{(e_1, e_1)}{\|(e_1, e_1)\|} = \frac{(e_1, e_1)}{\sqrt{2e_1^2}} = \left(\frac{e_1}{e_1\sqrt{2}}, \frac{e_1}{e_1\sqrt{2}}\right) = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

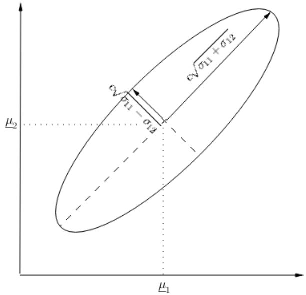
De manera análoga, se encuentra que el segundo vector-propio asociado al segundo valor-propio:

$\lambda_2 = \sigma_{11} - \sigma_{12}$ es:

$$\underline{\boldsymbol{e}}_2^t = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

Ahora, Como $\rho = \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} > 0$, entonces $\sigma_{12} > 0$ y por lo tanto $\lambda_1 > \lambda_2$, pues: $\quad \lambda_1 = \sigma_{11} + \sigma_{12} > \sigma_{11} > \sigma_{11} - \sigma_{12} = \lambda_2$.

De lo anterior se concluye que si $\sigma_{11} = \sigma_{22}$ y $\rho > 0$, entonces el eje mayor de la elipse está a lo largo de una línea cuya inclinación es de 45^o y que pasa por el punto $\underline{\mu} = (\mu_1, \mu_2)$.



Ejemplo-3: Suponga que el vector aleatorio $\underline{\boldsymbol{x}}^t = (X_1, X_2)$ tiene una distribución normal-bivariada, con vector de medias $\underline{\mu} = (\mu_1, \mu_2)$ y matriz de Var-Cov dada por $\boldsymbol{\Sigma}$. Halla la f.d.p multivariada de $\underline{\boldsymbol{x}}$.

Solución: Recordemos que la f.d.p normal bi-variada está dada por:

$$f(\underline{\boldsymbol{x}}) = (2\pi)^{-\frac{2}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\underline{\boldsymbol{x}}-\underline{\boldsymbol{\mu}})^t \boldsymbol{\Sigma}^{-1}(\underline{\boldsymbol{x}}-\underline{\boldsymbol{\mu}})}.$$

Para

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}, \qquad \boldsymbol{\Sigma}^{-1} = \frac{1}{|\Sigma|} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}$$

con $|\boldsymbol{\Sigma}| = \sigma_{11}\sigma_{22} - \sigma_{12}^2$, pero $\sigma_{12} = \rho\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}$, es decir que $|\boldsymbol{\Sigma}| = \sigma_{11}\sigma_{22} - \rho^2\sigma_{11}\sigma_{22} = \sigma_{11}\sigma_{22}(1-\rho^2)$, de donde:

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_{11}\sigma_{22}(1-\rho^2)} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}$$

Ahora se tiene que:

$$(\underline{\boldsymbol{x}}-\underline{\boldsymbol{\mu}})^t\boldsymbol{\Sigma}^{-1}(\underline{\boldsymbol{x}}-\underline{\boldsymbol{\mu}}) =$$

$$\frac{1}{\sigma_{11}\sigma_{22}(1-\rho^2)}\begin{bmatrix} x_1-\mu_1 & x_2-\mu_2 \end{bmatrix}\begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}$$

$$= \frac{1}{\sigma_{11}\sigma_{22}(1-\rho^2)}\sigma_{22}(x_1-\mu_1)^2+\sigma_{11}(x_2-\mu_2)^2 - 2\rho\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}(x_1-\mu_1)(x_2-\mu_2)$$

$$= \frac{1}{1-\rho^2}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_1-\mu_2}{\sqrt{\sigma_{22}}}\right)+\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2\right]$$

de donde

$$\frac{1}{2}(\underline{\boldsymbol{x}}-\underline{\boldsymbol{\mu}})^t\boldsymbol{\Sigma}^{-1}(\underline{\boldsymbol{x}}-\underline{\boldsymbol{\mu}}) =$$

$$\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_1-\mu_2}{\sqrt{\sigma_{22}}}\right)+\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2\right]$$

y como $|\boldsymbol{\Sigma}| = \sigma_{11}\sigma_{22}(1-\rho^2)$, entonces

$$|\boldsymbol{\Sigma}|^{-\frac{1}{2}} = \left[\sigma_{11}\sigma_{22}(1-\rho^2)\right]^{-\frac{1}{2}}$$

por lo tanto,

$$f(\underline{\boldsymbol{x}}) = f(x_1,x_2) = (2\pi)^{-\frac{2}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\underline{\boldsymbol{x}}-\underline{\boldsymbol{\mu}})^t\boldsymbol{\Sigma}^{-1}(\underline{\boldsymbol{x}}-\underline{\boldsymbol{\mu}})}$$

$$= (2\pi)^{-1}\left[\sigma_{11}\sigma_{22}(1-\rho^2)\right]^{-\frac{1}{2}}$$

$$\text{Exp}\left[-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_1-\mu_2}{\sqrt{\sigma_{22}}}\right) + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2\right]\right]$$

$$= (2\pi)^{-1}\left[\sigma_{11}\sigma_{22}(1-\rho^2)\right]^{-\frac{1}{2}}\text{Exp}\left[-\frac{1}{2(1-\rho^2)}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2\right]$$

$$\text{Exp}\left[-\frac{1}{2(1-\rho^2)}\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2\right]\text{Exp}\left[\frac{\rho}{1-\rho^2}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_1-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]$$

Ahora, si las variables X_1 y X_2 tienen $\rho = 0$ entonces:

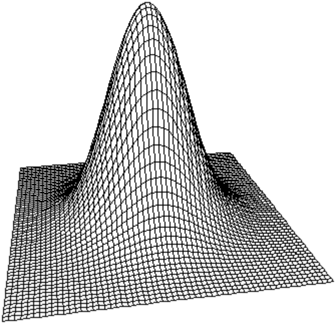
$$f(\underline{\boldsymbol{x}}) = f(x_1, x_2) = (2\pi)^{-1}|\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\underline{\boldsymbol{x}}-\underline{\boldsymbol{\mu}})^t \boldsymbol{\Sigma}^{-1}(\underline{\boldsymbol{x}}-\underline{\boldsymbol{\mu}})}$$

$$= (2\pi)^{-1} \left[\sigma_{11}\sigma_{22}\right]^{-\frac{1}{2}} \operatorname{Exp}\left[-\frac{1}{2}\left(\frac{x_1 - \mu_1}{\sqrt{\sigma_{11}}}\right)^2 - \frac{1}{2}\left(\frac{x_2 - \mu_2}{\sqrt{\sigma_{22}}}\right)^2\right]$$

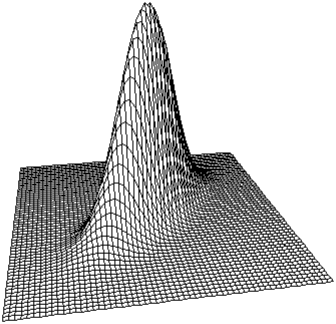
$$= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_{11}}} \operatorname{Exp}\left[-\frac{1}{2}\left(\frac{x_1 - \mu_1}{\sqrt{\sigma_{11}}}\right)^2\right] \frac{1}{\sqrt{2\pi}\sqrt{\sigma_{22}}} \operatorname{Exp}\left[-\frac{1}{2}\left(\frac{x_2 - \mu_2}{\sqrt{\sigma_{22}}}\right)^2\right]$$

$$f(x_1, x_2) = f(x_1) f(x_2)$$

con $f(x_1)$ la f.d.p de una normal uni-variada con media μ_1 y varianza σ_{11} y y $f(x_2)$ la f.d.p de una normal uni-variada con media μ_2 y varianza σ_{22}, ie.,si $\rho = 0$ entonces X_1 y X_2 son independientes, pues $f(x_1, x_2) = f(x_1)f(x_2)$.



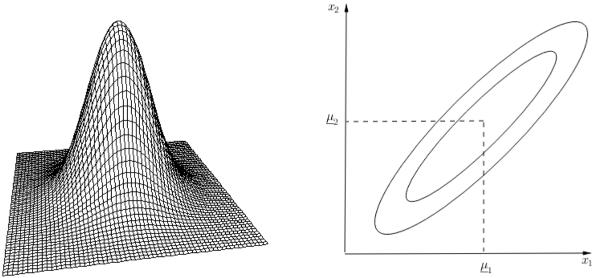
$\sigma_{11} = \sigma_{22}, \quad \rho = 0$



$\sigma_{11} = \sigma_{22}, \quad \rho = 0.750$

Notar que la presencia de correlación causa que la probabilidad se concentre a lo largo de una línea.

La densidad normal multivariada tiene un máximo valor cuando la distancia cuadrática, $(\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}})^t \boldsymbol{\Sigma}^{-1} (\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}})$ es igual a cero, es decir, cuando $\underline{\boldsymbol{x}} = \underline{\boldsymbol{\mu}}$. Por tanto el punto $\underline{\boldsymbol{\mu}}$ es el punto de máxima densidad, o la moda, y también es la media, como se observa en la siguiente gráfica.



Contornos de probabilidad

Definicion: Función Generadora de Momentos (FGM)

Caso Univariado:

Si una v.a $X \sim N_1\left(\mu\ ,\ \sigma^2\right)$, entonces la FGM de X es:

$$M_X(t) := E\left[e^{tX}\right] = e^{\mu t + \frac{1}{2}\sigma^2 t^2} = e^{t^t \mu + \frac{1}{2} t^t \sigma^2 t},\ \ t \in \boldsymbol{R}$$

Definicion: Función Generadora de Momentos (FGM)

Caso Multvariado:

Si un vector aleatorio $\underline{X} \sim N_p\left(\underline{\mu}\ ,\ \boldsymbol{\Sigma}\right)$, entonces la FGM de $\underline{X}$ es:

$$M_{\underline{X}}(\underline{t}) := E\left[e^{\underline{t}^t \underline{X}}\right] = e^{\underline{t}^t \underline{\mu} + \frac{1}{2}\underline{t}^t \boldsymbol{\Sigma} \underline{t}},\ \ \ \underline{t} \in \mathbb{R}^p$$

Propiedades de la distribución Normal Multivariada

1. Si $\underline{X} \sim N_p(\underline{\mu}, \mathbf{\Sigma})$, entonces

$$\underline{a}^t \underline{X} = a_1 X_1 + a_2 X_2 + \ldots + a_p X_p \sim N_1\left(\underline{a}^t \underline{\mu} \ , \ \underline{a}^t \mathbf{\Sigma} \underline{a}\right)$$

   Análogamente,

si $\forall \ \underline{a} \in \mathbb{R}^p$: $\quad \underline{a}^t \underline{X} \sim N_1\left(\underline{a}^t \underline{\mu} \ , \ \underline{a}^t \mathbf{\Sigma} \underline{a}\right)$ entonces $\underline{X} \sim N_p\left(\underline{\mu} \ , \ \mathbf{\Sigma}\right)$,

esdecir : $\quad \underline{X} \sim N_p\left(\underline{\mu} \ , \ \mathbf{\Sigma}\right) \Longleftrightarrow \underline{a}^t \underline{X} \sim N_1\left(\underline{a}^t \underline{\mu} \ , \ \underline{a}^t \mathbf{\Sigma} \underline{a}\right)$

Luego, si $\underline{X} \sim N_p(\underline{\mu} \ , \ \mathbf{\Sigma})$ entonces cada $X_i \sim N_1\left(\mu_i \ , \ \sigma_{ii}\right)$,

lo cual se logra con $\underline{a} = (0, 0, \cdots, 1, \cdots, 0, 0)^t$ con 1-en la posición i-ésima del vector $\underline{a}$ y $\underline{\mu}$ y $\mathbf{\Sigma}$ dados por:

$$\underline{\boldsymbol{\mu}} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \quad \text{y} \quad \mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

Demostración de 1:

$$\underline{X} \sim N_p(\underline{\mu}\ ,\ \mathbf{\Sigma}) \Longleftrightarrow \underline{a}^t \underline{X} \sim N_1(\underline{a}^t \underline{\mu}\ ,\ \underline{a}^t \mathbf{\Sigma} \underline{a})$$

Sea la v.a $Y = \underline{a}^t \underline{X}$

$$M_Y(t) = E[e^{tY}] = E\left[e^{t(\underline{a}^t \underline{X})}\right] = E\left[e^{(\underline{a}t)^t \underline{X}}\right] \text{, pues } (\underline{a}t)^t = t^t \underline{a}^t = t\underline{a}^t$$

$$= M_{\underline{X}}\left(\underline{a}t\right) = e^{(\underline{a}t)^t \underline{\mu} + \frac{1}{2}(\underline{a}t)^t \mathbf{\Sigma}(\underline{a}t)}$$

$$= e^{t^t \underline{a}^t \underline{\mu} + \frac{1}{2} t^t \underline{a}^t \mathbf{\Sigma} \underline{a} t} = e^{t(\underline{a}^t \underline{\mu}) + \frac{1}{2} t(\underline{a}^t \mathbf{\Sigma} \underline{a}) t}$$

$$= e^{t(\underline{a}^t \underline{\mu}) + \frac{1}{2}(\underline{a}^t \mathbf{\Sigma} \underline{a}) t^2}$$

es decir, la FGM de $Y = \underline{a}^t \underline{X}$-corresponde a la FGM de una v.a normal univariada con media $\underline{a}^t \underline{\mu}$ y varianza $\underline{a}^t \mathbf{\Sigma} \underline{a}$, es decir:

$$Y = \underline{a}^t \underline{X} \sim N_1\left(\underline{a}^t \underline{\mu}\ ,\ \underline{a}^t \mathbf{\Sigma} \underline{a}\right), \quad l.q.q.d$$

Ejemplo-4

Suponga que $\underline{\boldsymbol{x}} \sim N_3(\underline{\mu}\ ,\ \boldsymbol{\Sigma})$ y considere la combinación linea de las componentes de $\underline{\boldsymbol{x}}$ dada por:

$$\underline{\boldsymbol{y}} = 2X_1 - X_2 + 3X_3 = \begin{bmatrix} 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \underline{a}^t \underline{\boldsymbol{x}}\ , \text{ con } \underline{a} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}, \textit{ie.}$$

$$E[\underline{\boldsymbol{y}}] = E[\underline{a}^t \underline{\boldsymbol{x}}] = \underline{a}^t \underline{\mu} = \begin{bmatrix} 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = 2\mu_1 - \mu_2 + 3\mu_3\ , \quad \text{y}$$

$$Var[\underline{\boldsymbol{y}}] = Var[\underline{a}^t\underline{\boldsymbol{x}}] = \underline{a}^t\boldsymbol{\Sigma}\underline{a} = \begin{bmatrix} 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$$

$$= \begin{bmatrix} 2\sigma_{11} - \sigma_{12} + 3\sigma_{13} & 2\sigma_{12} - \sigma_{22} + 3\sigma_{23} & 2\sigma_{13} - \sigma_{23} + 3\sigma_{33} \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$$

$$= 4\sigma_{11} - 2\sigma_{12} + 6\sigma_{13} - 2\sigma_{12} + \sigma_{22} - 3\sigma_{23} + 6\sigma_{13} - 3\sigma_{23} + 9\sigma_{33}$$

$$Var[\underline{\boldsymbol{y}}] = 4\sigma_{11} + \sigma_{22} + 9\sigma_{33} - 4\sigma_{12} + 12\sigma_{13} - 6\sigma_{23}$$

Es decir que:

$$\underline{\boldsymbol{y}} = \underline{a}^t\underline{X} = 2X_1 - X_2 + 3X_3 \sim N_1\left(\underline{a}^t\underline{\mu}\ ,\ \underline{a}^t\boldsymbol{\Sigma}\underline{a}\right), \quad \text{con},$$

$$\underline{a}^t\underline{\mu} = 2\mu_1 - \mu_2 + 3\mu_3$$

$$\underline{a}^t\boldsymbol{\Sigma}\underline{a} = 4\sigma_{11} + \sigma_{22} + 9\sigma_{33} - 4\sigma_{12} + 12\sigma_{13} - 6\sigma_{23}$$

2. Si $\underline{\boldsymbol{x}} \sim N_p(\underline{\mu}\ ,\ \boldsymbol{\Sigma})$, entonces: $\underline{\boldsymbol{y}} = A\underline{X} \sim N_q(A\underline{\mu}\ ,\ A\boldsymbol{\Sigma}A^t)$

Demostración de 2:

$$M_{\underline{\boldsymbol{y}}}(\underline{t}) = E\left[e^{\underline{t}^t \underline{\boldsymbol{y}}}\right] = E\left[e^{\underline{t}^t (A\underline{\boldsymbol{x}})}\right] = E\left[e^{\underline{t}^t A\underline{\boldsymbol{x}}}\right]$$

$$= E\left[e^{(A^t\underline{t})^t \underline{\boldsymbol{x}}}\right]; \quad pues,\ \ (A^t\underline{t})^t = \underline{t}^t A$$

$$= M_{\underline{X}}\left(A^t\underline{t}\right) = e^{(A^t\underline{t})^t\underline{\mu} + \frac{1}{2}(A^t\underline{t})^t\boldsymbol{\Sigma}(A^t\underline{t})};$$

$$pues,\ \ M_{\underline{X}}(\underline{t}) = e^{\underline{t}^t\underline{\mu} + \frac{1}{2}\underline{t}^t\boldsymbol{\Sigma}\underline{t}}$$

$$= e^{\underline{t}^t A\underline{\mu} + \frac{1}{2}\underline{t}^t A\boldsymbol{\Sigma}A^t\underline{t}}$$

lo cual corresponde a la FGM de un Vector aleatorio normal multivariado con vector de medias $A\underline{\mu}$ y matriz de varianzas-covarainzas $A\boldsymbol{\Sigma}A^t$, es decir:

$$\underline{\boldsymbol{y}} = A\underline{X} \sim N_q\Big(A\underline{\mu}\ ,\ A\boldsymbol{\Sigma}A^t\Big)\ ,\quad l.q.q.d.$$

Ejemplo-5

Suponga que $\underline{\boldsymbol{x}} \sim N_3(\underline{\mu}\ ,\ \boldsymbol{\Sigma})$ y considere el vector de combinaciones lineales de las variables de $\underline{\boldsymbol{x}}$ dado por:

$$\underline{\boldsymbol{z}} = A\underline{\boldsymbol{x}} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} X_1 - X_2 \\ X_2 - X_3 \end{bmatrix} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$$

entonces,

$$\underline{\boldsymbol{z}} = A\underline{X} \sim N_q\Big(A\underline{\mu}\ ,\ A\boldsymbol{\Sigma}A^t\Big), \quad \text{donde}$$

$$A\underline{\mu} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_2 - \mu_3 \end{bmatrix} = \begin{bmatrix} E[Z_1] \\ \\ E[Z_2] \end{bmatrix}$$

$$A\boldsymbol{\Sigma}A^t = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} - \sigma_{12} & \sigma_{12} - \sigma_{22} & \sigma_{13} - \sigma_{23} \\ \\ \sigma_{12} - \sigma_{13} & \sigma_{22} - \sigma_{23} & \sigma_{23} - \sigma_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} - \sigma_{12} - \sigma_{12} + \sigma_{22} & \sigma_{12} - \sigma_{22} - \sigma_{13} + \sigma_{23} \\ \\ \sigma_{12} - \sigma_{13} - \sigma_{22} + \sigma_{23} & \sigma_{22} - \sigma_{23} - \sigma_{23} + \sigma_{33} \end{bmatrix}$$

$$A\boldsymbol{\Sigma}A^t = \begin{bmatrix} \sigma_{11} - 2\sigma_{12} + \sigma_{22} & \sigma_{12} - \sigma_{22} - \sigma_{13} + \sigma_{23} \\ \\ \sigma_{12} - \sigma_{22} - \sigma_{13} + \sigma_{23} & \sigma_{22} - 2\sigma_{23} + \sigma_{33} \end{bmatrix}$$

Ejemplo-6

Suponga que $\underline{\boldsymbol{x}}' = (X_1, X_2, X_3) \sim N_3(\underline{\mu}\ ,\ \boldsymbol{\Sigma})$, donde:

$$\underline{\mu} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} \quad , \quad \text{y} \quad \Sigma = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

Hallar la distribución conjunta de probabilidad de:

$Z_1 = X_1 + X_2 + X_3$ y $Z_2 = X_1 - X_2$.

Sea

$$\underline{\boldsymbol{z}} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} X_1 + X_2 + X_3 \\ X_1 - X_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = A\underline{\boldsymbol{x}}, \quad \text{luego,}$$

$$\underline{\boldsymbol{z}} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \sim N_2\Big(A\underline{\mu}\ ,\ A\boldsymbol{\Sigma}A^t\Big),$$

donde:

$$A\underline{\mu} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}_{2\times 1}$$

y

$$A\mathbf{\Sigma}A^t = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 4 & 4 & 2 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 10 & 0 \\ 0 & 3 \end{bmatrix}$$

de lo anterior, Z_1 y Z_2 son independientes, pues $Cov(Z_1, Z_2) = 0$ y Z_1 y Z_2 son distribuidas normales uni-variadas.

3. Sea $\underline{X} \sim N_p(\underline{\mu}\ ,\ \boldsymbol{\Sigma})$ y sean los vectores y matrices particionadas dadas por:

$$\underline{\boldsymbol{x}} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} \underline{\boldsymbol{x}}^{(1)} \\ \cdots \\ \underline{\boldsymbol{x}}^{(2)} \end{bmatrix} \ ,\ \underline{\mu} = \begin{bmatrix} \underline{\mu}^{(1)} \\ \cdots \\ \underline{\mu}^{(2)} \end{bmatrix} \text{ y } \boldsymbol{\Sigma}_{p\times p} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & | & \boldsymbol{\Sigma}_{12} \\ -- & -- & -- \\ \boldsymbol{\Sigma}_{21} & | & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

entonces:

$$\underline{\boldsymbol{x}}^{(1)} \sim N_q\left(\underline{\mu}^{(1)}\ ,\ \boldsymbol{\Sigma}_{11}\right)$$

$$\underline{\boldsymbol{x}}^{(2)} \sim N_{p-q}\left(\underline{\mu}^{(2)}\ ,\ \boldsymbol{\Sigma}_{22}\right).$$

La anterior propiedad también se puede enunciar como sigue:

Todos los subconjuntos de variables de $\underline{\boldsymbol{x}}$ tienen distribución normal, sea univariada o multivariada.

Demostración de 3:

Para esto se usará la siguiente matriz particionada:

$$A_{q\times p} = \begin{array}{cc} & \begin{array}{ccc} q & & p-q \end{array} \\ q & \left[\begin{array}{ccc} \boldsymbol{I}_q & \vdots & \boldsymbol{O} \end{array}\right]_{q\times p} \end{array} = \left[\begin{array}{ccc} \boldsymbol{I}_q & \vdots & \boldsymbol{O} \end{array}\right]_{q\times p}, \quad \text{luego,}$$

$$A\underline{X} = \left[\begin{array}{ccc} \boldsymbol{I}_q & \vdots & \boldsymbol{O} \end{array}\right]_{q\times p} \begin{bmatrix} \underline{\boldsymbol{x}}^{(1)} \\ \cdots \\ \underline{\boldsymbol{x}}^{(2)} \end{bmatrix}_{p\times 1}$$

$$= \left[\begin{array}{ccc} \boldsymbol{I}_q\ \underline{\boldsymbol{x}}^{(1)} & + & \boldsymbol{O}\ \underline{\boldsymbol{x}}^{(2)} \end{array}\right]_{q\times 1}$$

$$= \left[\begin{array}{c} \boldsymbol{I}_q\ \underline{\boldsymbol{x}}^{(1)} \end{array}\right]_{q\times 1}$$

$$A\underline{X} = \left[\begin{array}{c} \underline{\boldsymbol{x}}^{(1)} \end{array}\right]_{q\times 1} = \underline{\boldsymbol{x}}^{(1)}$$

de donde, usando la propiedad (2) se tiene que:

$$A\underline{X} = \underline{\boldsymbol{x}}^{(1)} \sim N_q(A\underline{\mu}\ ,\ A\boldsymbol{\Sigma}A^t)$$

Pero,

$$A\underline{\mu} = \begin{bmatrix} \boldsymbol{I}_q & \vdots & \boldsymbol{O} \end{bmatrix} \begin{bmatrix} \underline{\mu}^{(1)} \\ \cdots \\ \underline{\mu}^{(2)} \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{I}_q\ \underline{\mu}^{(1)} & + & \boldsymbol{O}\ \underline{\mu}^{(2)} \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{I}_q\ \underline{\mu}^{(1)} \end{bmatrix}$$

$$A\underline{\mu} = \begin{bmatrix} \underline{\mu}^{(1)} \end{bmatrix}_{q\times 1} = \underline{\mu}^{(1)}$$

$$A\boldsymbol{\Sigma}A^t = \begin{bmatrix} \boldsymbol{I}_q & \vdots & \boldsymbol{O} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{11} & | & \boldsymbol{\Sigma}_{12} \\ -- & -- & -- \\ \boldsymbol{\Sigma}_{21} & | & \boldsymbol{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_q \\ \cdots \\ \boldsymbol{O} \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{I}_q\ \boldsymbol{\Sigma}_{11} + \boldsymbol{O}\ \boldsymbol{\Sigma}_{21} & \vdots & \boldsymbol{I}_q\ \boldsymbol{\Sigma}_{12} + \boldsymbol{O}\ \boldsymbol{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_q \\ \cdots \\ \boldsymbol{O} \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{\Sigma}_{11} + \boldsymbol{O} & \vdots & \boldsymbol{\Sigma}_{12} + \boldsymbol{O} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_q \\ \cdots \\ \boldsymbol{O} \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \vdots & \boldsymbol{\Sigma}_{12} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_q \\ \cdots \\ \boldsymbol{O} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{11}\boldsymbol{I}_q + \boldsymbol{\Sigma}_{12}\boldsymbol{O} \end{bmatrix} = \boldsymbol{\Sigma}_{11}$$

Es decir que:

$$\underline{\boldsymbol{x}}^{(1)} \sim N_q\Big(A\underline{\mu}\ ,\ A\boldsymbol{\Sigma}A^t\Big)$$

$$N_q\Big(\underline{\mu}^{(1)}\ ,\ \boldsymbol{\Sigma}_{11}\Big)\ , \qquad l.q.q.d.$$

De manera similar, se demuestra que: (TAREA):

$$\underline{\boldsymbol{x}}^{(2)} \sim N_q\Big(\underline{\mu}^{(2)}\ ,\ \boldsymbol{\Sigma}_{22}\Big),$$

en este caso usando la matriz particionada dada por:

$$A = \begin{array}{cc} & \begin{array}{ccc} q & & p-q \end{array} \\ p-q & \left[\begin{array}{ccc} \boldsymbol{O} & \vdots & \boldsymbol{I}_{p-q} \end{array}\right] \end{array} = \left[\begin{array}{ccc} \boldsymbol{O} & \vdots & \boldsymbol{I}_{p-q} \end{array}\right]$$

Ejemplo-7 Sea $\underline{\boldsymbol{x}} \sim N_5(\underline{\mu}\ ,\ \boldsymbol{\Sigma})$, hallar la distribución de: $\begin{bmatrix} X_2 \\ X_4 \end{bmatrix}$

$$\text{Sean,} \quad \underline{\boldsymbol{x}}^{(1)} = \begin{bmatrix} X_2 \\ X_4 \end{bmatrix} \ , \quad \underline{\boldsymbol{\mu}}^{(1)} = \begin{bmatrix} \mu_2 \\ \mu_4 \end{bmatrix} \quad \text{y} \quad \boldsymbol{\Sigma}_{11} = \begin{bmatrix} \sigma_{22} & \sigma_{24} \\ \sigma_{24} & \sigma_{44} \end{bmatrix}$$

Asumiendo que $\underline{\boldsymbol{x}}$, $\underline{\boldsymbol{\mu}}$ y $\boldsymbol{\Sigma}$ son particionados como sigue:

$$\underline{\boldsymbol{x}} = \begin{bmatrix} X_2 \\ X_4 \\ \cdots \\ X_1 \\ X_3 \\ X_5 \end{bmatrix} \ , \quad \underline{\boldsymbol{\mu}} = \begin{bmatrix} \mu_2 \\ \mu_4 \\ \cdots \\ \mu_1 \\ \mu_3 \\ \mu_5 \end{bmatrix} \quad \text{y} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{22} & \sigma_{24} & \vdots & \sigma_{12} & \sigma_{23} & \sigma_{25} \\ \sigma_{24} & \sigma_{44} & \vdots & \sigma_{14} & \sigma_{34} & \sigma_{45} \\ \cdots & \cdots\cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_{12} & \sigma_{14} & \vdots & \sigma_{11} & \sigma_{13} & \sigma_{15} \\ \sigma_{23} & \sigma_{34} & \vdots & \sigma_{13} & \sigma_{33} & \sigma_{35} \\ \sigma_{25} & \sigma_{45} & \vdots & \sigma_{15} & \sigma_{35} & \sigma_{55} \end{bmatrix}$$

$$\text{es decir :} \quad \underline{\boldsymbol{x}} = \begin{bmatrix} \underline{\boldsymbol{x}}^{(1)} \\ \cdots \\ \underline{\boldsymbol{x}}^{(2)} \end{bmatrix} \quad , \quad \underline{\mu} = \begin{bmatrix} \underline{\mu}^{(1)} \\ \cdots \\ \underline{\mu}^{(2)} \end{bmatrix} \quad \text{y} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & | & \boldsymbol{\Sigma}_{12} \\ -- & -- & -- \\ \boldsymbol{\Sigma}_{21} & | & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

$$\text{luego :} \quad \underline{\boldsymbol{x}}^{(1)} \sim N_2\left(\underline{\mu}^{(1)} \, , \, \boldsymbol{\Sigma}_{11}\right) \sim N_2\left(\begin{bmatrix} \mu_2 \\ \mu_4 \end{bmatrix} \, , \, \begin{bmatrix} \sigma_{22} & \sigma_{24} \\ \sigma_{24} & \sigma_{44} \end{bmatrix} \right) .$$

4. a). Si $\underline{\boldsymbol{x}}^{(1)}$ y $\underline{\boldsymbol{x}}^{(2)}$ son estadísticamente independientes, entonces $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}^t = O$.

b).

$$\text{Si,} \quad \underline{\boldsymbol{x}} = \begin{bmatrix} \underline{\boldsymbol{x}}^{(1)} \\ \cdots \\ \underline{\boldsymbol{x}}^{(2)} \end{bmatrix} \sim N_p \left[\begin{pmatrix} \underline{\mu}^{(1)} \\ \cdots \\ \underline{\mu}^{(2)} \end{pmatrix} , \begin{pmatrix} \boldsymbol{\Sigma}_{11} & | & \boldsymbol{\Sigma}_{12} \\ -- & -- & -- \\ \boldsymbol{\Sigma}_{21} & | & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right]$$

entonces,

$$\underline{\boldsymbol{x}}^{(1)} \perp \underline{\boldsymbol{x}}^{(2)} \quad Sii \quad \boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}^t = O.$$

Ejemplo Suponga que $\underline{\boldsymbol{x}} \sim N_3(\underline{\mu} \,,\, \boldsymbol{\Sigma})$, con

$$\boldsymbol{\Sigma} = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

¿Son las variables X_1 y X_2 independientes?

Rta. NO, porque $\sigma_{12} = 1 \neq 0$.

¿Son los siguientes vectores aleatorios independientes?:

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} X_3 \end{bmatrix}$$

$$\underline{\boldsymbol{x}} = \begin{bmatrix} X_1 \\ X_2 \\ \cdots \\ X_3 \end{bmatrix} \ , \ \underline{\boldsymbol{\mu}} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdots \\ \mu_3 \end{bmatrix} \quad \text{y} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \vdots & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \vdots & \sigma_{23} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{31} & \sigma_{32} & \vdots & \sigma_{33} \end{bmatrix} = \begin{bmatrix} 4 & 1 & \vdots & 0 \\ 1 & 3 & \vdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \vdots & 2 \end{bmatrix}$$

$$\text{luego para :} \quad \underline{\boldsymbol{x}}^{(1)} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \quad \text{y} \quad \underline{\boldsymbol{x}}^{(2)} = \begin{bmatrix} X_3 \end{bmatrix}$$

$$\text{se tiene:} \quad \boldsymbol{\Sigma}_{12} = Cov\left(\underline{\boldsymbol{x}}^{(1)} \ , \ \underline{\boldsymbol{x}}^{(2)}\right) = \begin{bmatrix} \sigma_{13} \\ \sigma_{23} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

por lo tanto, si son independientes.

c) Si $\underline{\boldsymbol{x}}^{(1)} \perp \underline{\boldsymbol{x}}^{(2)}$ y

$$\underline{\boldsymbol{x}}^{(1)} \sim N_q\left(\underline{\mu}^{(1)}\ ,\ \boldsymbol{\Sigma}_{11}\right) \quad \text{y} \quad \underline{\boldsymbol{x}}^{(2)} \sim N_q\left(\underline{\mu}^{(2)}\ ,\ \boldsymbol{\Sigma}_{22}\right)$$

entonces,

$$\underline{\boldsymbol{x}} = \begin{bmatrix} \underline{\boldsymbol{x}}^{(1)} \\ \cdots \\ \underline{\boldsymbol{x}}^{(2)} \end{bmatrix} \sim N_p\left[\begin{pmatrix} \underline{\mu}^{(1)} \\ \cdots \\ \underline{\mu}^{(2)} \end{pmatrix}\ ,\ \begin{pmatrix} \boldsymbol{\Sigma}_{11} & | & O \\ -- & -- & -- \\ O & | & \boldsymbol{\Sigma}_{22} \end{pmatrix}\right]$$

Además:

$$\underline{\boldsymbol{x}}^{(1)} + \underline{\boldsymbol{x}}^{(2)} \sim N_k\left(\underline{\mu}^{(1)} + \underline{\mu}^{(2)}\ ,\ \boldsymbol{\Sigma}_{11} + \boldsymbol{\Sigma}_{22}\right), \quad \text{si} \quad q = p - q = k$$

Demostración de 4(c): Para esto se usarán la FGM de un vector normal-multivariado. Sean

$$\underline{\boldsymbol{x}}^{(1)} \perp \underline{\boldsymbol{x}}^{(2)} \ , \ \underline{\boldsymbol{x}}^{(1)} \sim N_q\left(\underline{\mu}^{(1)} \ , \ \boldsymbol{\Sigma}_{11}\right) \ \text{ y } \ \underline{\boldsymbol{x}}^{(2)} \sim N_q\left(\underline{\mu}^{(2)} \ , \ \boldsymbol{\Sigma}_{22}\right)$$

$$M_{\underline{\boldsymbol{x}}^{(1)}+\underline{\boldsymbol{x}}^{(2)}}(\underline{t}) := E\left[e^{\underline{t}^t\left(\underline{\boldsymbol{x}}^{(1)}+\underline{\boldsymbol{x}}^{(2)}\right)}\right] = E\left[e^{\underline{t}^t\underline{\boldsymbol{x}}^{(1)}}e^{\underline{t}^t\underline{\boldsymbol{x}}^{(2)})}\right]$$

$$= E\left[e^{\underline{t}^t\underline{\boldsymbol{x}}^{(1)}}\right]E\left[e^{\underline{t}^t\underline{\boldsymbol{x}}^{(2)})}\right], \quad \text{pues, } \underline{\boldsymbol{x}}^{(1)} \perp \underline{\boldsymbol{x}}^{(2)}$$

$$= M_{\underline{\boldsymbol{x}}^{(1)}}(\underline{t})M_{\underline{\boldsymbol{x}}^{(2)}}(\underline{t}) = e^{\underline{t}^t\underline{\mu}^{(1)}+\frac{1}{2}\underline{t}^t\boldsymbol{\Sigma}_{11}\underline{t}}e^{\underline{t}^t\underline{\mu}^{(2)}+\frac{1}{2}\underline{t}^t\boldsymbol{\Sigma}_{22}\underline{t}}$$

$$= e^{\underline{t}^t\underline{\mu}^{(1)}+\underline{t}^t\underline{\mu}^{(2)}+\frac{1}{2}\underline{t}^t\boldsymbol{\Sigma}_{11}\underline{t}+\frac{1}{2}\underline{t}^t\boldsymbol{\Sigma}_{22}\underline{t}} = e^{\underline{t}^t\left(\underline{\mu}^{(1)}+\underline{\mu}^{(2)}\right)+\frac{1}{2}\underline{t}^t\left(\boldsymbol{\Sigma}_{11}+\boldsymbol{\Sigma}_{22}\right)\underline{t}}$$

lo cual corresponde a la FGM de un vector aleatorio normal multivariado con vector de medias $\underline{\mu}^{(1)} + \underline{\mu}^{(2)}$ y matriz de varianzas covarianzas $\mathbf{\Sigma}_{11} + \mathbf{\Sigma}_{22}$,

es decir:

$$\underline{\boldsymbol{x}}^{(1)} + \underline{\boldsymbol{x}}^{(2)} \sim N_k\left(\underline{\mu}^{(1)} + \underline{\mu}^{(2)}\ ,\ \mathbf{\Sigma}_{11} + \mathbf{\Sigma}_{22}\right)\ , \quad l.q.q.d$$

Demostración de 4(a) y 4(b); (Ejercicio 4.14 del libro de Johnson, sixth edition)).

5. Si

$$\underline{\boldsymbol{x}} = \begin{bmatrix} \underline{\boldsymbol{x}}^{(1)} \\ \cdots \\ \underline{\boldsymbol{x}}^{(2)} \end{bmatrix} \sim N_p \left[\begin{pmatrix} \underline{\mu}^{(1)} \\ \cdots \\ \underline{\mu}^{(2)} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & | & \boldsymbol{\Sigma}_{12} \\ -- & -- & -- \\ \boldsymbol{\Sigma}_{21} & | & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right] \quad ; \; |\boldsymbol{\Sigma}_{22}| > 0$$

entonces, la distribución condicional de $\underline{\boldsymbol{x}}^{(1)}$ dado $\underline{\boldsymbol{x}}^{(2)}$ esta dada por:

$$\underline{\boldsymbol{x}}^{(1)} | \underline{\boldsymbol{x}}^{(2)} = \underline{\boldsymbol{x}}^{(2)} \sim N_q \left[\underline{\mu}^{(1)} + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \left(\underline{\boldsymbol{x}}^{(2)} - \underline{\mu}^{(2)} \right), \; \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \right]$$

es decir:

$$E \left[\underline{\boldsymbol{x}}^{(1)} | \underline{\boldsymbol{x}}^{(2)} = \underline{\boldsymbol{x}}^{(2)} \right] = \underline{\mu}^{(1)} + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \left(\underline{\boldsymbol{x}}^{(2)} - \underline{\mu}^{(2)} \right)$$

$$Var \left[\underline{\boldsymbol{x}}^{(1)} | \underline{\boldsymbol{x}}^{(2)} = \underline{\boldsymbol{x}}^{(2)}] \right] = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}$$

De manera similar, si $|\boldsymbol{\Sigma}_{11}| > 0$, entonces:

$$\underline{\boldsymbol{x}}^{(2)} | \underline{\boldsymbol{x}}^{(1)} = \underline{\boldsymbol{x}}^{(1)} \sim N_{p-q} \left[\underline{\mu}^{(2)} + \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \left(\underline{\boldsymbol{x}}^{(1)} - \underline{\mu}^{(1)} \right), \; \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \right]$$

Demostración de (5): Considere la matriz $p \times p$ dada por:

$$\boldsymbol{A} = \left[\begin{array}{c|c} \boldsymbol{I}_{q\times q} & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}{}_{q\times(p-q)} \\ \hline \boldsymbol{0}_{(p-q)\times q} & \boldsymbol{I}_{(p-q)\times(p-q)} \end{array}\right]_{p\times p}$$

teniendo en cuenta que el siguiente vector aleatorio:

$$(\underline{X} - \underline{\mu}) \sim N_p(\underline{0}\ ,\ \boldsymbol{\Sigma})$$

se tiene que:

$$\underline{Z}_{p\times 1} = \boldsymbol{A}(\underline{X} - \underline{\mu}) = \left[\begin{array}{c|c} \boldsymbol{I} & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \hline \boldsymbol{0} & \boldsymbol{I} \end{array}\right] \begin{bmatrix} \underline{X}^{(1)} - \underline{\mu}^{(1)} \\ \underline{X}^{(2)} - \underline{\mu}^{(2)} \end{bmatrix}$$

$$= \begin{bmatrix} \underline{X}^{(1)} - \underline{\mu}^{(1)} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left(\underline{X}^{(2)} - \underline{\mu}^{(2)}\right) \\ \underline{X}^{(2)} - \underline{\mu}^{(2)} \end{bmatrix} \sim N_p(\underline{\mu}_{\underline{Z}}\ ,\ \boldsymbol{\Sigma}_{\underline{Z}})$$

con, $\quad \underline{\mu}_{\underline{Z}} = \boldsymbol{A}\underline{0} = \underline{0} \quad y \quad \boldsymbol{\Sigma}_{\underline{Z}} = \boldsymbol{A\Sigma A}^t$

$$\boldsymbol{A\Sigma A}^t = \left[\begin{array}{c|c} \boldsymbol{I} & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \hline \boldsymbol{0} & \boldsymbol{I} \end{array}\right] \left[\begin{array}{c|c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right] \boldsymbol{A}^t$$

$$= \left[\begin{array}{c|c} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \boldsymbol{0} \\ \hline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right] \left[\begin{array}{c|c} \boldsymbol{I} & \boldsymbol{0} \\ \hline -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \boldsymbol{I} \end{array}\right]$$

$$\boldsymbol{\Sigma}_{\underline{Z}} = \left[\begin{array}{c|c} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \boldsymbol{0} \\ \hline \boldsymbol{0} & \boldsymbol{\Sigma}_{22} \end{array}\right],$$ es decir que:

$$\underline{\boldsymbol{Z}} = \begin{bmatrix} \underline{X}^{(1)} - \underline{\mu}^{(1)} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left(\underline{X}^{(2)} - \underline{\mu}^{(2)}\right) \\ \\ \underline{X}^{(2)} - \underline{\mu}^{(2)} \end{bmatrix} \sim N_p\left(\underline{\mu}_{\underline{Z}}\,,\ \boldsymbol{\Sigma}_{\underline{Z}}\right) = N_p\left(\underline{0}\,,\ \boldsymbol{\Sigma}_{\underline{Z}}\right)$$

de lo anterior se tiene que como los vectores aleatorios:

$$\underline{Z}_1 = \underline{X}^{(1)} - \underline{\mu}^{(1)} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\left(\underline{X}^{(2)} - \underline{\mu}^{(2)}\right) \quad \text{y} \quad \underline{Z}_2 = \underline{X}^{(2)} - \underline{\mu}^{(2)}$$

tienen matriz de var-cov $\mathbf{0}$, entonces son independientes (por (4.b)), y además se cumple que:

$$\underline{Z}_1 = \underline{X}^{(1)} - \underline{\mu}^{(1)} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\left[\underline{X}^{(2)} - \underline{\mu}^{(2)}\right] \sim N_p\left(\underline{\mathbf{0}}\ ,\ \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\right)$$

por lo tanto, la distribución condicional de:

$$\underline{Z}_1 = \underline{X}^{(1)} - \underline{\mu}^{(1)} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\left(\underline{X}^{(2)} - \underline{\mu}^{(2)}\right) \quad \text{dado} \quad \underline{Z}_2 = \underline{X}^{(2)} - \underline{\mu}^{(2)}$$

es igual a la marginal de:

$$\underline{Z}_1 = \underline{X}^{(1)} - \underline{\mu}^{(1)} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\left(\underline{X}^{(2)} - \underline{\mu}^{(2)}\right)$$

cuando $\underline{X}^{(2)}$ toma un valor fijo, es decir,

$$\underline{\boldsymbol{x}}^{(1)} \mid \underline{\boldsymbol{x}}^{(2)} = \underline{X}^{(2)} \sim N_q\left[\underline{\mu}^{(1)} + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\left(\underline{X}^{(2)} - \underline{\mu}^{(2)}\right),\ \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\right]$$

Ejemplo-9 Distribución condicional de una Normal-Bivariada.

Sea $\underline{\boldsymbol{x}} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N_2 \left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \;,\; \begin{pmatrix} \sigma_{11} & | & \sigma_{12} \\ -- & -- & -- \\ \sigma_{21} & | & \sigma_{22} \end{pmatrix} \right]$, con $\sigma_{22} > 0$,

luego, la distribución condicional de $X_1 \big| X_2$ esta dada por:

$$X_1 \mid X_2 = x_2 \sim N_1 \left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2) \;,\; \sigma_{11} - \frac{\sigma_{12}}{\sigma_{22}}\sigma_{21} \right)$$

pues,

$$\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} = \sigma_{12}\sigma_{22}^{-1} = \frac{\sigma_{12}}{\sigma_{22}} \;, \quad \text{y}$$

$$\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} = \sigma_{11} - \sigma_{12}\sigma_{22}^{-1}\sigma_{21} = \sigma_{11} - \frac{\sigma_{12}}{\sigma_{22}}\sigma_{21}$$

y como, $\quad \rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}}$, es decir: $\sqrt{\sigma_{11}}\rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{22}}}$, $\quad$ luego,

$$\frac{\sigma_{12}}{\sigma_{22}} = \frac{\sqrt{\sigma_{11}}}{\sqrt{\sigma_{22}}}\rho_{12} \quad \text{y} \quad \frac{\sigma_{12}^2}{\sigma_{22}} = \frac{\sigma_{11}\sigma_{22}\rho_{12}^2}{\sigma_{22}} = \sigma_{11}\rho_{12}^2$$

y

$$\sigma_{11} - \frac{\sigma_{12}}{\sigma_{22}}\sigma_{21} = \sigma_{11}\left[1 - \rho_{12}^2\right]$$

es decir:

$$X_1 \mid X_2 = x_2 \sim N_1\left(\mu_1 + \frac{\sqrt{\sigma_{11}}}{\sqrt{\sigma_{22}}}\rho_{12}(x_2 - \mu_2)\ ,\ \sigma_{11}\left[1 - \rho_{12}^2\right]\right)$$

$$\sim N_1\left(\underbrace{\mu_1 - \frac{\sqrt{\sigma_{11}}}{\sqrt{\sigma_{22}}}\rho_{12}\mu_2}_{a_0} + \underbrace{\frac{\sqrt{\sigma_{11}}}{\sqrt{\sigma_{22}}}\rho_{12}}_{b}x_2\ ,\ \sigma_{11}\left[1 - \rho_{12}^2\right]\right)$$

$$X_1 \mid X_2 = x_2 \sim N_1\left(a_0 + bx_2\ ,\ \sigma_{11}\left[1 - \rho_{12}^2\right]\right)$$

de donde se observa que la media de la distribución condicional de $X_1 \mid X_2 = x_2$, corresponde a la ecuación de una línea recta con intecepto a_0 y pendiente b, ie. la ecuación del modelo de regresión lineal de X_1 v.s X_2.

Observaciones:

a) En regresión multivariada, la media condicional

$$\underline{\mu}_{1.2} = E\left[\underline{X}^{(1)} \mid \underline{X}^{(2)}\right]$$

es llamada la curva de regresión.

b) Es decir, la curva de regresión en la normal multivariada,

$$\underline{\mu}_{1.2} = E\left[\underline{X}^{(1)} \mid \underline{X}^{(2)}\right],$$

se puede escribir como:

$$E\left[\underline{X}^{(1)} \mid \underline{X}^{(2)}\right] = \begin{bmatrix} E[X_1|X_{q+1}, X_{q+2}, \cdots, X_p] \\ \\ E[X_2|X_{q+1}, X_{q+2}, \cdots, X_p] \\ \vdots \qquad\qquad\qquad \vdots \\ \vdots \qquad\qquad\qquad \vdots \\ E[X_q|X_{q+1}, X_{q+2}, \cdots, X_p] \end{bmatrix} = \underline{\mu}^{(1)} + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\left(\underline{X}^{(2)} - \underline{\mu}^{(2)}\right)$$

Ahora con, $\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1} = \begin{bmatrix} \beta_{1,q+1} & \beta_{1,q+2} & \cdots & \beta_{1,p} \\ \\ \beta_{2,q+1} & \beta_{2,q+2} & \cdots & \beta_{2,p} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \beta_{q,q+1} & \beta_{q,q+2} & \cdots & \beta_{q,p} \end{bmatrix}$

entonces, la curva de regresión se puede escribir como:

$$E[\underline{X}^{(1)} \mid \underline{X}^{(2)}] = \underline{\mu}^{(1)} + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\left(\underline{X}^{(2)} - \underline{\mu}^{(2)}\right)$$

$$= \begin{bmatrix} E[X_1|X_{q+1}, X_{q+2}, \cdots, X_p] \\ E[X_2|X_{q+1}, X_{q+2}, \cdots, X_p] \\ \vdots \\ \vdots \\ E[X_q|X_{q+1}, X_{q+2}, \cdots, X_p] \end{bmatrix}$$

$$= \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \vdots \\ \mu_q \end{bmatrix} + \begin{bmatrix} \beta_{1,q+1} & \beta_{1,q+2} & \cdots & \beta_{1,p} \\ \beta_{2,q+1} & \beta_{2,q+2} & \cdots & \beta_{2,p} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \beta_{q,q+1} & \beta_{q,q+2} & \cdots & \beta_{q,p} \end{bmatrix} \begin{bmatrix} X_{q+1} - \mu_{q+1} \\ X_{q+2} - \mu_{q+2} \\ \vdots \\ \vdots \\ X_p - \mu_p \end{bmatrix}$$

$$E\left[\underline{X}^{(1)} \mid \underline{X}^{(2)}\right] = \begin{bmatrix} \mu_1 + \beta_{1,q+1}(X_{q+1} - \mu_{q+1}) + \cdots\cdots + \beta_{1,p}(X_p - \mu_p) \\ \\ \mu_2 + \beta_{2,q+1}(X_{q+1} - \mu_{q+1}) + \cdots\cdots + \beta_{2,p}(X_p - \mu_p) \\ \vdots \qquad\qquad\qquad \vdots \\ \vdots \qquad\qquad\qquad \vdots \\ \mu_q + \beta_{q,q+1}(X_{q+1} - \mu_{q+1}) + \cdots\cdots + \beta_{q,p}(X_p - \mu_p) \end{bmatrix}$$

lo anterior implica que, cuando la distribución conjunta de las variables en una regresión (dependientes e independientes) es normal multivariada, todas las curvas de regresión son lineales.

c) La matriz de var-cov condicional

$$\mathbf{\Sigma}_{1.2} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$$

es constante pues no depende de los valores de las variables condicionantes. Por tanto, la curva de regresión es homocedástica.

Ejemplo-10 Suponga que $\underline{\boldsymbol{x}}' = (X_1, X_2, X_3) \sim N_p(\underline{\mu}\ ,\ \boldsymbol{\Sigma})$, donde:

$$\underline{\mu} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} \ , \quad \text{y} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

Hallar la distribución condicional de: $\underline{\boldsymbol{x}}^{(1)} \mid \underline{\boldsymbol{x}}^{(2)}$, donde:

$$\underline{\boldsymbol{x}}^{(1)} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \quad \text{y} \quad \underline{\boldsymbol{x}}^{(2)} = \begin{bmatrix} X_3 \end{bmatrix}$$

La distribución condicional de: $\underline{\boldsymbol{x}}^{(1)} \mid \underline{\boldsymbol{x}}^{(2)}$ es normal bi-variada con vector de medias:

$$E\left[\underline{\boldsymbol{x}}^{(1)} \mid \underline{\boldsymbol{x}}^{(2)}\right] = \underline{\mu}^{(1)} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left(\underline{\boldsymbol{x}}^{(2)} - \underline{\mu}^{(2)}\right)$$

y matriz de var-cov dada por:

$$Var\left[\underline{\boldsymbol{x}}^{(1)} \mid \underline{\boldsymbol{x}}^{(2)}\right] = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$$

Haciendo,

$$\underline{\mu} = \begin{bmatrix} \underline{\mu}^{(1)} \\ \cdots \\ \underline{\mu}^{(2)} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdots \\ \mu_3 \end{bmatrix} = \begin{bmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} \\ \cdots \\ \begin{pmatrix} 2 \end{pmatrix} \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & | & \boldsymbol{\Sigma}_{12} \\ -- & -- & -- \\ \boldsymbol{\Sigma}_{21} & | & \boldsymbol{\Sigma}_{22} \end{pmatrix} = \begin{bmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} & \vdots & \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ \cdots & & \cdots \\ \begin{pmatrix} 1 & 0 \end{pmatrix} & \vdots & \begin{bmatrix} 1 \end{bmatrix} \end{bmatrix}$$

luego,

$$E\left[\underline{\boldsymbol{x}}^{(1)} \mid \underline{\boldsymbol{x}}^{(2)}\right] = \underline{\mu}^{(1)} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left(\underline{\boldsymbol{x}}^{(2)} - \underline{\mu}^{(2)}\right) = \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \frac{1}{1} \left(x_3 - 2\right) = \begin{pmatrix} x_3 \\ 1 \end{pmatrix}$$

y

$$Var\left[\underline{\boldsymbol{x}}^{(1)} \mid \underline{\boldsymbol{x}}^{(2)}\right] = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} \frac{1}{1} \begin{pmatrix} 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 3 \end{pmatrix}$$

6. Sea $\underline{X} \sim N_p(\underline{\mu}\ ,\ \mathbf{\Sigma})$. Si $|\mathbf{\Sigma}| > 0$, entonces:

$$\underline{\boldsymbol{z}} = \mathbf{\Sigma}^{-1/2}(\underline{X} - \underline{\mu}) \sim N_p\Big(\underline{0}\ ,\ \boldsymbol{I}_p\Big),$$

ie, $\underline{\boldsymbol{z}}$-tiene una distribución normal-multivariada estándar.

Demostración de la propiedad (6): Estudiarla del libro de Jonhson.

7. Sea $\underline{X} \sim N_p(\underline{\mu}\ ,\ \mathbf{\Sigma})$, con $|\mathbf{\Sigma}| > 0$.

a) $(\underline{X} - \underline{\mu})^t \mathbf{\Sigma}^{-1} (\underline{X} - \underline{\mu}) \sim \chi^2_p$

b) La distribución $N_p(\underline{\mu}\ ,\ \mathbf{\Sigma})$-asigna probabilidad de $(1-\alpha)100\%$ al elipsoide determinado por:

$$\left\{\underline{\boldsymbol{x}}\ :\ (\underline{\boldsymbol{x}} - \underline{\mu})^t \mathbf{\Sigma}^{-1} (\underline{\boldsymbol{x}} - \underline{\mu}) \leq \chi^2_{\alpha;p}\right\}$$

donde, $\chi^2_{\alpha;p}$-es el percentil superior α de la distribución χ^2_p.

Demostración: (a).

Sea

$$\underline{Z} = \mathbf{\Sigma}^{-1/2}(\underline{X} - \underline{\mu}),$$

donde $\mathbf{\Sigma}^{-1/2}$ es la matriz inversa de $\mathbf{\Sigma}^{1/2} = \mathbf{\Gamma}\Delta^{1/2}\mathbf{\Gamma}^t$ (llamada la matriz raíz cuadrada positiva de la matriz $\mathbf{\Sigma}$).

entonces

$$\underline{Z} = \mathbf{\Sigma}^{-1/2}(\underline{X} - \underline{\mu}) \sim N_p(\underline{0}\ ,\ \boldsymbol{I}_p) \qquad (Resultado - 6)$$

luego, Las marginales de las variables del vector $\underline{Z}$ son $N(0,1)$ e independientes.

Ahora, se Considera la variable:

$$(\underline{X} - \underline{\mu})^t \mathbf{\Sigma}^{-1}(\underline{X} - \underline{\mu}) = (\underline{X} - \underline{\mu})^t \mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{-1/2}(\underline{X} - \underline{\mu})$$

$$= \left(\mathbf{\Sigma}^{-1/2}(\underline{X} - \underline{\mu})\right)^t \left(\mathbf{\Sigma}^{-1/2}(\underline{X} - \underline{\mu})\right) = \underline{Z}^t \underline{Z} = \sum_{i=1}^{p} Z_i^2 \sim \chi^2_{(p)}$$

Observaciones:

- Por el teorema de descomposición espectral de la matriz simétrica y definida positiva $\boldsymbol{\Sigma}$ se tiene que:

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Delta}\boldsymbol{\Gamma}^t$$

donde $\boldsymbol{\Gamma}$ es una matriz ortogonal con los vectores propios de $\boldsymbol{\Sigma}$ como columnas ($\boldsymbol{\Gamma}^T\boldsymbol{\Gamma} = \boldsymbol{I}$), y $\boldsymbol{\Delta}$ es una matriz diagonal con los valores propios de $\boldsymbol{\Sigma}$ en su diagonal. Entonces,

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Delta}\boldsymbol{\Gamma}^t = \boldsymbol{\Gamma}\boldsymbol{\Delta}^{1/2}\boldsymbol{\Delta}^{1/2}\boldsymbol{\Gamma}^t = \left(\boldsymbol{\Gamma}\boldsymbol{\Delta}^{1/2}\boldsymbol{\Gamma}^t\right)\left(\boldsymbol{\Gamma}\boldsymbol{\Delta}^{1/2}\boldsymbol{\Gamma}^t\right) = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2}$$

donde $\boldsymbol{\Delta}^{1/2}$ es una matriz diagonal con la raíz cuadrada de los elementos de la diagonal de $\boldsymbol{\Delta}$ en su diagonal.

A $\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Gamma}\boldsymbol{\Delta}^{1/2}\boldsymbol{\Gamma}^t$ se le llama la matriz raíz cuadrada positiva de la matriz $\boldsymbol{\Sigma}$.

- El empleo de la matriz $\boldsymbol{\Sigma}^{-1/2}$ sobre el vector aleatorio $(\underline{X} - \underline{\mu})$, estandariza todas las variables y elimina los efectos de correlación entre ellas.

8. Sean $\underline{\boldsymbol{x}}_1, \underline{\boldsymbol{x}}_2, \ldots, \underline{\boldsymbol{x}}_n$, vectores aleatorios p-variados mutuamente-independientes tales que:

$$\underline{\boldsymbol{x}}_i \sim N_p\Big(\underline{\mu}_i \ , \ \boldsymbol{\Sigma}\Big) \ , \quad \text{para} \quad i = 1, 2, \ldots, n,$$

entonces:

$$\underline{\boldsymbol{v}}_1 = \sum_{i=1}^{n} c_i \underline{\boldsymbol{x}}_i = c_1 \underline{\boldsymbol{x}}_1 + c_2 \underline{\boldsymbol{x}}_2 + \cdots + c_n \underline{\boldsymbol{x}}_n$$

$$\sim N_p \left[\sum_{i=1}^{n} c_i \underline{\mu}_i \ , \ \left(\sum^{n} c_i^2 \right) \boldsymbol{\Sigma} \right]$$

$$\underline{\boldsymbol{v}}_1 \sim N_p\Big(c_1 \underline{\mu}_1 + c_2 \underline{\mu}_2 + \cdots + c_n \underline{\mu}_n \ , \ [c_1^2 + c_2^2 + \cdots + c_n^2] \boldsymbol{\Sigma} \Big)$$

además si, $\quad \underline{\boldsymbol{v}}_2 = \sum_{i=1}^{n} b_i \underline{\boldsymbol{x}}_i = b_1 \underline{\boldsymbol{x}}_1 + b_2 \underline{\boldsymbol{x}}_2 + \cdots + b_n \underline{\boldsymbol{x}}_n$

entonces: $\quad \underline{\boldsymbol{v}} = \begin{bmatrix} \underline{\boldsymbol{v}}_1 \\ \cdots \\ \underline{\boldsymbol{v}}_2 \end{bmatrix} \sim N_{2p} \left[\begin{pmatrix} \sum_{i=1}^{n} c_i \underline{\boldsymbol{\mu}}_i \\ \cdots\cdots \\ \sum_{i=1}^{n} b_i \underline{\boldsymbol{\mu}}_i \end{pmatrix} , \ \boldsymbol{\Sigma}_{\underline{\boldsymbol{v}}} \right]$, donde

$$\boldsymbol{\Sigma}_{\underline{\boldsymbol{v}}} = Var(\underline{\boldsymbol{v}}) = \begin{bmatrix} \left(\sum_{i=1}^{n} c_i^2\right) \boldsymbol{\Sigma} & \vdots & \left(\sum_{i=1}^{n} b_i c_i\right) \boldsymbol{\Sigma} \\ \cdots\cdots & & \cdots\cdots \\ \left(\sum_{i=1}^{n} b_i c_i\right) \boldsymbol{\Sigma} & \vdots & \left(\sum_{i=1}^{n} b_i^2\right) \boldsymbol{\Sigma} \end{bmatrix}$$

$$= \begin{bmatrix} (\underline{c}^t \underline{c}) \boldsymbol{\Sigma} & \vdots & (\underline{b}^t \underline{c}) \boldsymbol{\Sigma} \\ \cdots\cdots & & \cdots\cdots \\ (\underline{b}^t \underline{c}) \boldsymbol{\Sigma} & \vdots & (\underline{b}^t \underline{b}) \boldsymbol{\Sigma} \end{bmatrix}$$

y además, $\quad \underline{\boldsymbol{v}}_1 \perp \underline{\boldsymbol{v}}_2 \quad$ si $\quad \underline{b}^t \underline{c} = \sum_{i=1}^{n} b_i c_i = 0.$

Demostración:

Considere el vector $np \times 1$ dado por:

$$\underline{Z} = \begin{bmatrix} \underline{\boldsymbol{x}}_1 \\ \underline{\boldsymbol{x}}_2 \\ \vdots \\ \vdots \\ \underline{\boldsymbol{x}}_n \end{bmatrix}_{np\times 1} \quad , \quad \text{recuerde que:} \quad \underset{p\times 1}{\underline{\boldsymbol{x}}_i}$$

por el resultado (4.c) se tiene que:

$$\underline{Z} \sim N_{np}\left(\underline{\mu}_{\underline{Z}} \, , \, \boldsymbol{\Sigma}_{\underline{Z}}\right) \quad \text{donde:}$$

$$\underline{\mu}_{\underline{Z}} = \begin{bmatrix} \underline{\mu}_1 \\ \underline{\mu}_2 \\ \vdots \\ \vdots \\ \underline{\mu}_n \end{bmatrix}_{np\times 1} \qquad \text{y} \qquad \boldsymbol{\Sigma}_{\underline{Z}} = \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma} \end{bmatrix}_{np\times np}$$

1. para la parte (a):

Considere la matriz

$$\boldsymbol{A} = \begin{bmatrix} c_1\boldsymbol{I} & | & c_2\boldsymbol{I} & | & \cdots & | & c_n\boldsymbol{I} \end{bmatrix}_{p\times np}$$

donde $\boldsymbol{I}$ es la matriz identidad $p \times p$, entonces

$$\boldsymbol{A}\underline{Z} = \begin{bmatrix} c_1\boldsymbol{I} & | & c_2\boldsymbol{I} & | & \cdots & | & c_n\boldsymbol{I} \end{bmatrix} \begin{bmatrix} \underline{\boldsymbol{x}}_1 \\ \underline{\boldsymbol{x}}_2 \\ \vdots \\ \vdots \\ \underline{\boldsymbol{x}}_n \end{bmatrix} = c_1\underline{\boldsymbol{x}}_1 + c_2\underline{\boldsymbol{x}}_2 + \cdots + c_n\underline{\boldsymbol{x}}_n = \underline{\boldsymbol{v}}_1$$

y por el resultado o propiedad (2) se tiene que:

$$\boldsymbol{A}\underline{Z} = \underline{\boldsymbol{v}}_1 \sim N_p\left(\boldsymbol{A}\underline{\mu}_{\underline{Z}}\ ,\ \boldsymbol{A}\boldsymbol{\Sigma}_{\underline{Z}}\boldsymbol{A}^t\right) \quad \text{donde:}$$

$$\boldsymbol{A}\underline{\mu}_{\underline{Z}} = \begin{bmatrix} c_1\boldsymbol{I} & | & c_2\boldsymbol{I} & | & \cdots & | & c_n\boldsymbol{I} \end{bmatrix} \begin{bmatrix} \underline{\mu}_1 \\ \underline{\mu}_2 \\ \vdots \\ \vdots \\ \underline{\mu}_n \end{bmatrix} = \textstyle\sum_{i=1}^n c_i\underline{\mu}_i, \quad y$$

$$\boldsymbol{A}\boldsymbol{\Sigma}_{\underline{Z}}\boldsymbol{A}^t = \begin{bmatrix} c_1\boldsymbol{I} & | & c_2\boldsymbol{I} & | & \cdots & | & c_n\boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{\Sigma} \end{bmatrix} \begin{bmatrix} c_1\boldsymbol{I} \\ c_2\boldsymbol{I} \\ \vdots \\ \vdots \\ c_n\boldsymbol{I} \end{bmatrix}$$

$$= \begin{bmatrix} c_1\boldsymbol{\Sigma} & c_2\boldsymbol{\Sigma} & \cdots & c_n\boldsymbol{\Sigma} \end{bmatrix} \begin{bmatrix} c_1\boldsymbol{I} \\ c_2\boldsymbol{I} \\ \vdots \\ \vdots \\ c_n\boldsymbol{I} \end{bmatrix} = \left(\sum_{i=1}^n c_i^2\right)\boldsymbol{\Sigma}\ , \quad \text{es decir,}$$

$$\underline{\boldsymbol{v}}_1 = \sum_{i=1}^{n} c_i \underline{X}_i \sim N_p \left[\sum_{i=1}^{n} c_i \underline{\mu}_i \ , \ \left(\sum_{i=1}^{n} c_i^2 \right) \boldsymbol{\Sigma} \right] \ , \quad l.q.q.d$$

2. para la parte (b):

Sea

$$\boldsymbol{B} = \begin{bmatrix} c_1\boldsymbol{I} & | & c_2\boldsymbol{I} & | & \cdots & | & c_n\boldsymbol{I} \\ & & & & & & \\ b_1\boldsymbol{I} & | & b_2\boldsymbol{I} & | & \cdots & | & b_n\boldsymbol{I} \end{bmatrix}_{2p\times np}$$

entonces,

$$\underline{\boldsymbol{v}} = \begin{bmatrix} \underline{\boldsymbol{v}}_1 \\ \\ \underline{\boldsymbol{v}}_2 \end{bmatrix} = \begin{bmatrix} c_1\underline{\boldsymbol{x}}_1 + c_2\underline{\boldsymbol{x}}_2 + \cdots + c_n\underline{\boldsymbol{x}}_n \\ \\ b_1\underline{\boldsymbol{x}}_1 + b_2\underline{\boldsymbol{x}}_2 + \cdots + b_n\underline{\boldsymbol{x}}_n \end{bmatrix}$$

$$= \begin{bmatrix} c_1\boldsymbol{I} & | & c_2\boldsymbol{I} & | & \cdots & | & c_n\boldsymbol{I} \\ & & & & & & \\ b_1\boldsymbol{I} & | & b_2\boldsymbol{I} & | & \cdots & | & b_n\boldsymbol{I} \end{bmatrix} \begin{bmatrix} \underline{\boldsymbol{x}}_1 \\ \underline{\boldsymbol{x}}_2 \\ \vdots \\ \vdots \\ \underline{\boldsymbol{x}}_n \end{bmatrix} = \boldsymbol{B}\underline{Z}$$

y del resultado (2) nuevamente se tiene que:

$$\boldsymbol{B}\underline{Z} \sim N_{2p}\left(\boldsymbol{B}\underline{\mu}_{\underline{Z}}\ ,\ \boldsymbol{B\Sigma}_{\underline{Z}}\boldsymbol{B}^t\right)$$

donde,

$$\boldsymbol{B}\underline{\mu}_{\underline{Z}} = \begin{bmatrix} c_1\boldsymbol{I} & | & c_2\boldsymbol{I} & | & \cdots & | & c_n\boldsymbol{I} \\ \\ b_1\boldsymbol{I} & | & b_2\boldsymbol{I} & | & \cdots & | & b_n\boldsymbol{I} \end{bmatrix} \begin{bmatrix} \underline{\mu}_1 \\ \underline{\mu}_2 \\ \vdots \\ \vdots \\ \underline{\mu}_n \end{bmatrix} = \begin{bmatrix} \sum\limits_{=1}^{n} c_i \underline{\boldsymbol{\mu}}_i \\ ------ \\ \sum\limits_{=1}^{n} b_i \underline{\boldsymbol{\mu}}_i \end{bmatrix}$$

$$\boldsymbol{B\Sigma}_{\underline{Z}}\boldsymbol{B}^t = \begin{bmatrix} c_1\boldsymbol{I} & | & c_2\boldsymbol{I} & | & \cdots & | & c_n\boldsymbol{I} \\ \\ b_1\boldsymbol{I} & | & b_2\boldsymbol{I} & | & \cdots & | & b_n\boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{\Sigma} \end{bmatrix} \begin{bmatrix} c_1\boldsymbol{I} & b_1\boldsymbol{I} \\ c_2\boldsymbol{I} & b_2\boldsymbol{I} \\ \vdots & \\ \vdots & \\ c_n\boldsymbol{I} & b_n\boldsymbol{I} \end{bmatrix}$$

$$\mathbf{\Sigma}_{\underline{\boldsymbol{v}}_1,\underline{\boldsymbol{v}}_2} = \begin{bmatrix} c_1\mathbf{\Sigma} & | & c_2\mathbf{\Sigma} & | & \cdots & | & c_n\mathbf{\Sigma} \\ \\ b_1\mathbf{\Sigma} & | & b_2\mathbf{\Sigma} & | & \cdots & | & b_n\mathbf{\Sigma} \end{bmatrix} \begin{bmatrix} c_1\boldsymbol{I} & b_1\boldsymbol{I} \\ c_2\boldsymbol{I} & b_2\boldsymbol{I} \\ \vdots & \\ \vdots & \\ c_n\boldsymbol{I} & b_n\boldsymbol{I} \end{bmatrix}$$

$$= \begin{bmatrix} \left(\sum_{i=1}^n c_i^2\right)\mathbf{\Sigma} & \vdots & \left(\sum_{i=1}^n b_i c_i\right)\mathbf{\Sigma} \\ \cdots\cdots & & \cdots\cdots \\ \left(\sum_{i=1}^n b_i c_i\right)\mathbf{\Sigma} & \vdots & \left(\sum_{i=1}^n b_i^2\right)\mathbf{\Sigma} \end{bmatrix} = \begin{bmatrix} \left(\underline{c}^t\underline{c}\right)\mathbf{\Sigma} & \vdots & \left(\underline{b}^t\underline{c}\right)\mathbf{\Sigma} \\ \cdots\cdots & & \cdots\cdots \\ \left(\underline{b}^t\underline{c}\right)\mathbf{\Sigma} & \vdots & \left(\underline{b}^t\underline{b}\right)\mathbf{\Sigma} \end{bmatrix}_{2p\times 2p}$$

es decir que,

$$\underline{\boldsymbol{v}} = \begin{bmatrix} \underline{\boldsymbol{v}}_1 \\ \cdots \\ \underline{\boldsymbol{v}}_2 \end{bmatrix} \sim N_{2p}\left[\begin{pmatrix} \sum_{i=1}^n c_i\underline{\boldsymbol{\mu}}_i \\ \cdots\cdots \\ \sum_{i=1}^n b_i\underline{\boldsymbol{\mu}}_i \end{pmatrix}, \; \mathbf{\Sigma}_{\underline{\boldsymbol{v}}}\right]$$

3. para la parte (c): Evidente por resultado (4.b).

Ejemplo-11 Suponga que $\underline{\boldsymbol{x}}_1, \underline{\boldsymbol{x}}_2, \underline{\boldsymbol{x}}_3, \underline{\boldsymbol{x}}_4$, son vectores aleatorios 3-variados independientes e idénticamente distribuidos, ie. $\underline{\boldsymbol{x}}_i = (X_{i1}, X_{i2}, X_{i3})^t$ para $i = 1, 2, 3, 4$, con:

$$\underline{\mu} = E[\underline{\boldsymbol{x}}_i] = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} \quad , \quad \text{y} \quad \boldsymbol{\Sigma} = Var[\underline{\boldsymbol{x}}_i] = \begin{bmatrix} 3 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix}$$

1. Hallar la media y varianza de: $\underline{a}^t \underline{\boldsymbol{x}}_1 = a_1 X_{11} + a_2 X_{12} + a_3 X_{13}$
2. Hallar la media y varianza de: $\frac{1}{2}\underline{\boldsymbol{x}}_1 + \frac{1}{2}\underline{\boldsymbol{x}}_2 + \frac{1}{2}\underline{\boldsymbol{x}}_3 + \frac{1}{2}\underline{\boldsymbol{x}}_4$
3. Hallar la media y varianza de:

$$\begin{bmatrix} \frac{1}{2}\underline{\boldsymbol{x}}_1 + \frac{1}{2}\underline{\boldsymbol{x}}_2 + \frac{1}{2}\underline{\boldsymbol{x}}_3 + \frac{1}{2}\underline{\boldsymbol{x}}_4 \\ \\ \underline{\boldsymbol{x}}_1 + \underline{\boldsymbol{x}}_2 + \underline{\boldsymbol{x}}_3 - 3\underline{\boldsymbol{x}}_4 \end{bmatrix}$$

Solución:

1. La distribución de:

$$\underline{a}^t \underline{\boldsymbol{x}}_1 = a_1 X_{11} + a_2 X_{12} + a_3 X_{13}.$$

La cual es una combinación lineal de las componentes de un vector aleatorio, ie. una c.l de variables, y por lo tanto es una variable aleatoria, de donde se tiene que:

$$\begin{aligned} E[\underline{a}^t \underline{\boldsymbol{x}}_1] &= E\Big[a_1 X_{11} + a_2 X_{12} + a_3 X_{13}\Big] \\ &= a_1 E[X_{11}] + a_2 E[X_{12}] + a_3 E[X_{13}] \\ &= 3a_1 - a_2 + a_3 \end{aligned}$$

$$Var[\underline{a}^t \underline{\mathbf{x}}_1] = Var\Big[a_1 X_{11} + a_2 X_{12} + a_3 X_{13}\Big]$$

$$= a_1^2 Var[X_{11}] + a_2^2 Var[X_{12}] + a_3^2 Var[X_{13}]$$

$$+ 2a_1 a_2 Cov(X_{11}, X_{12}) + 2a_1 a_3 Cov(X_{11}, X_{13}) + 2a_2 a_3 Cov(X_{12}, X_{13})$$

$$= 3a_1^2 + a_2^2 + 2a_3^2 - 2a_1 a_2 + 2a_1 a_3$$

$$Var[\underline{a}^t \underline{\mathbf{x}}_1] = \underline{a}^t \mathbf{\Sigma} \underline{a}.$$

2. La media y varianza de:

$$\frac{1}{2}\underline{\boldsymbol{x}}_1 + \frac{1}{2}\underline{\boldsymbol{x}}_2 + \frac{1}{2}\underline{\boldsymbol{x}}_3 + \frac{1}{2}\underline{\boldsymbol{x}}_4$$

En este caso se tiene un combinación lineal de vectores aleatorios, lo cual a su vez es un vector aleatorio, por lo tanto:

$$\begin{aligned}
E\left[\frac{1}{2}\underline{\boldsymbol{x}}_1 + \frac{1}{2}\underline{\boldsymbol{x}}_2 + \frac{1}{2}\underline{\boldsymbol{x}}_3 + \frac{1}{2}\underline{\boldsymbol{x}}_4\right] &= \frac{1}{2}E[\underline{\boldsymbol{x}}_1] + \frac{1}{2}E[\underline{\boldsymbol{x}}_2] + \frac{1}{2}E[\underline{\boldsymbol{x}}_3] + c_4 E[\underline{\boldsymbol{x}}_4] \\
&= \sum_{i=1}^{n} \frac{1}{2}\underline{\mu}_i = \frac{1}{2}\underline{\mu} + \frac{1}{2}\underline{\mu} + \frac{1}{2}\underline{\mu} + \frac{1}{2}\underline{\mu} \\
&= \left(\frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2}\right)\underline{\mu} \\
&= 2\underline{\mu} = 2\begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ -2 \\ 2 \end{bmatrix}
\end{aligned}$$

$$Var\left[c_1\underline{\boldsymbol{x}}_1 + c_2\underline{\boldsymbol{x}}_2 + c_3\underline{\boldsymbol{x}}_3 + c_4\underline{\boldsymbol{x}}_4\right]$$

$$= c_1^2 Var[\underline{\boldsymbol{x}}_1] + c_2^2 Var[\underline{\boldsymbol{x}}_2] + c_3^2 Var[\underline{\boldsymbol{x}}_3] + c_4^2 Var[\underline{\boldsymbol{x}}_4]$$

$$= c_1^2\boldsymbol{\Sigma} + c_2^2\boldsymbol{\Sigma} + c_3^2\boldsymbol{\Sigma} + c_4^2\boldsymbol{\Sigma} = \left(\sum_{i=1}^{n} c_i^2\right)\boldsymbol{\Sigma}$$

$$= \left(c_1^2 + c_2^2 + c_3^2 + c_4^2\right)\boldsymbol{\Sigma}$$

$$= \left(\frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4}\right)\boldsymbol{\Sigma} = \boldsymbol{\Sigma} = \begin{bmatrix} 3 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix}$$

3. La media y varianza de:

$$\begin{bmatrix} c_1\underline{\boldsymbol{x}}_1 + c_2\underline{\boldsymbol{x}}_2 + c_3\underline{\boldsymbol{x}}_3 + c_4\underline{\boldsymbol{x}}_4 \\ \\ b_1\underline{\boldsymbol{x}}_1 + b_2\underline{\boldsymbol{x}}_2 + b_3\underline{\boldsymbol{x}}_3 + b_4\underline{\boldsymbol{x}}_4 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\underline{\boldsymbol{x}}_1 + \frac{1}{2}\underline{\boldsymbol{x}}_2 + \frac{1}{2}\underline{\boldsymbol{x}}_3 + \frac{1}{2}\underline{\boldsymbol{x}}_4 \\ \\ \underline{\boldsymbol{x}}_1 + \underline{\boldsymbol{x}}_2 + \underline{\boldsymbol{x}}_3 - 3\underline{\boldsymbol{x}}_4 \end{bmatrix} = \begin{bmatrix} \underline{\boldsymbol{v}}_1 \\ \\ \underline{\boldsymbol{v}}_2 \end{bmatrix}_{6\times 1}$$

Sea $\underline{\boldsymbol{v}} = \begin{bmatrix} \underline{\boldsymbol{v}}_1 \\ \underline{\boldsymbol{v}}_2 \end{bmatrix}$, luego, $Var[\underline{\boldsymbol{v}}]$

$$= \begin{pmatrix} Var(\underline{\boldsymbol{v}}_1) & | & Cov(\underline{\boldsymbol{v}}_1 \ , \ \underline{\boldsymbol{v}}_2) \\ -- & -- & -- \\ Cov(\underline{\boldsymbol{v}}_2 \ , \ \underline{\boldsymbol{v}}_1) & | & Var(\underline{\boldsymbol{v}}_2) \end{pmatrix} = \begin{bmatrix} \left(\sum_{i=1}^n c_i^2\right)\boldsymbol{\Sigma} & \vdots & \left(\sum_{i=1}^n b_i c_i\right)\boldsymbol{\Sigma} \\ \cdots\cdots & & \cdots\cdots \\ \left(\sum_{i=1}^n b_i c_i\right)\boldsymbol{\Sigma} & \vdots & \left(\sum_{i=1}^n b_i^2\right)\boldsymbol{\Sigma} \end{bmatrix}$$

$$= \begin{bmatrix} (\underline{c}^t\underline{c})\,\boldsymbol{\Sigma} & \vdots & (\underline{b}^t\underline{c})\,\boldsymbol{\Sigma} \\ \cdots\cdots & & \cdots\cdots \\ (\underline{b}^t\underline{c})\,\boldsymbol{\Sigma} & \vdots & (\underline{b}^t\underline{b})\,\boldsymbol{\Sigma} \end{bmatrix} = \begin{bmatrix} 1\times\boldsymbol{\Sigma} & \vdots & 0\times\boldsymbol{\Sigma} \\ \cdots\cdots & & \cdots\cdots \\ 0\times\boldsymbol{\Sigma} & \vdots & 12\times\boldsymbol{\Sigma} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma} & \vdots & \boldsymbol{0} \\ \cdots\cdots & & \cdots\cdots \\ \boldsymbol{0} & \vdots & 12\boldsymbol{\Sigma} \end{bmatrix}_{6\times 6}$$

es decir,

$$Var[\underline{\boldsymbol{v}}] = \left[\begin{array}{ccc} \boldsymbol{\Sigma} & \vdots & \mathbf{0} \\ \cdots\cdots & & \cdots\cdots \\ \mathbf{0} & \vdots & 12\boldsymbol{\Sigma} \end{array}\right] = \left[\begin{array}{ccccccc} 3 & -1 & 1 & & 0 & 0 & 0 \\ -1 & 1 & 0 & \vdots & 0 & 0 & 0 \\ 1 & 0 & 2 & & 0 & 0 & 0 \\ & \cdots\cdots & & & & \cdots\cdots & \\ 0 & 0 & 0 & & 36 & -12 & 12 \\ 0 & 0 & 0 & \vdots & -12 & 12 & 0 \\ 0 & 0 & 0 & & 12 & 0 & 24 \end{array}\right]$$

Observaciones:

- Cada componente de la primera combinación lineal es independiente de cada componente de la segunda combinación lineal.
- Conjuntamente las dos combinaciones lineales tienen una distribución normal multivariada 6 dimensional.
- Las dos combinaciones lineales son independientes.

$$\underline{\boldsymbol{v}} = \begin{bmatrix} \underline{\boldsymbol{v}}_1 \\ \\ \underline{\boldsymbol{v}}_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\underline{\boldsymbol{x}}_1 + \frac{1}{2}\underline{\boldsymbol{x}}_2 + \frac{1}{2}\underline{\boldsymbol{x}}_3 + \frac{1}{2}\underline{\boldsymbol{x}}_4 \\ \\ \underline{\boldsymbol{x}}_1 + \underline{\boldsymbol{x}}_2 + \underline{\boldsymbol{x}}_3 - 3\underline{\boldsymbol{x}}_4 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}(\underline{\boldsymbol{x}}_1 + \underline{\boldsymbol{x}}_2 + \underline{\boldsymbol{x}}_3 + \underline{\boldsymbol{x}}_4) \\ \\ \underline{\boldsymbol{x}}_1 + \underline{\boldsymbol{x}}_2 + \underline{\boldsymbol{x}}_3 - 3\underline{\boldsymbol{x}}_4 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{2}\begin{pmatrix} X_{11} \\ X_{12} \\ X_{13} \end{pmatrix} + \frac{1}{2}\begin{pmatrix} X_{21} \\ X_{22} \\ X_{23} \end{pmatrix} + \frac{1}{2}\begin{pmatrix} X_{31} \\ X_{32} \\ X_{33} \end{pmatrix} + \frac{1}{2}\begin{pmatrix} X_{41} \\ X_{42} \\ X_{43} \end{pmatrix} \\ \\ \begin{pmatrix} X_{11} \\ X_{12} \\ X_{13} \end{pmatrix} + \begin{pmatrix} X_{11} \\ X_{12} \\ X_{13} \end{pmatrix} + \begin{pmatrix} X_{11} \\ X_{12} \\ X_{13} \end{pmatrix} - 3\begin{pmatrix} X_{11} \\ X_{12} \\ X_{13} \end{pmatrix} \end{bmatrix}$$

$$
\underline{\boldsymbol{v}} = \begin{bmatrix} \underline{\boldsymbol{v}}_1 \\ \\ \underline{\boldsymbol{v}}_2 \end{bmatrix}_{6\times 1} = \begin{bmatrix} \frac{1}{2}X_{11} + \frac{1}{2}X_{21} + \frac{1}{2}X_{31} + \frac{1}{2}X_{41} \\ \\ \frac{1}{2}X_{12} + \frac{1}{2}X_{22} + \frac{1}{2}X_{32} + \frac{1}{2}X_{42} \\ \\ \frac{1}{2}X_{13} + \frac{1}{2}X_{23} + \frac{1}{2}X_{33} + \frac{1}{2}X_{43} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ X_{11} + X_{21} + X_{31} - 3X_{41} \\ X_{12} + X_{22} + X_{32} - 3X_{42} \\ X_{13} + X_{23} + X_{33} - 3X_{43} \end{bmatrix}
$$

$$
\boldsymbol{\Sigma}_{\underline{\boldsymbol{x}}_i} = \begin{bmatrix} 3 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix} \quad , \quad \boldsymbol{\Sigma}_{\underline{\boldsymbol{v}}} = \left[\begin{array}{ccc|ccc} 3 & -1 & 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 2 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 36 & -12 & 12 \\ 0 & 0 & 0 & -12 & 12 & 0 \\ 0 & 0 & 0 & 12 & 0 & 24 \end{array}\right]
$$

Evaluación del Supuesto de Normalidad Multivariado

Evaluar el supuesto de Normalidad Multivariada es importante para facilitar los procesos de inferencia estadística.

Existen varios métodos o alternativas para evaluar el supuesto de Normalidad Multivariada de un vector aleatorio.

Cuando n-es grande y los métodos de evaluación utilizados se basan en el vector de medias muestrales $\underline{\overline{\boldsymbol{x}}}$ o en ciertas distancias que involucran dicho vector de medias muestrales, el supuesto de normalidad multivariada parece no ser tan crucial.

En este caso la calidad de las inferencias realizadas, dependerá de qué tan parecida sea la distribución del vector de medias muestrales $\underline{\overline{\boldsymbol{x}}}$ a una normal multivariada.

También es importante tener métodos para identificar cuando la distribución de un vector aleatorio se aleja de la normal multivariada, y así tener cuidado con los análisis posteriores.

Propiedad de la Normal-Multivariada:

Recordar que bajo el supuesto de normalidad multivariada de un vector aleatorio $\underline{\boldsymbol{x}}_{p\times 1}$, cualquier combinación lineal de las componentes de dicho vector tiene una distribución normal univariada.

Los gráficos de contorno de la normal tri-variada son elipsoides (o hiperelipsoides) y para el caso particular de $p = 2$ (normal bi-variada) son elipses.

Preguntas Adecuadas:

Algunos pasos previos antes de evaluar la normalidad multivariada, corresponden a responder las siguientes preguntas:

1. ¿Son las distribuciones marginales del vector aleatorio, parecídas a normales univariadas?
2. ¿Es la distribución de alguna combinación lineal de las componentes de $\underline{\boldsymbol{x}}_{p\times 1}$, NO parecida a una normal univariada?
3. ¿Al hacer gráficos de dispersión por pares de componentes de $\underline{\boldsymbol{x}}_{p\times 1}$, presentan algunos de ellos comportamientos no elípticos?
4. ¿Existen datos atípicos a nivel marginal o a nivel bivariado?

Métodos prácticos:

En la práctica para evaluar el supuesto de normalidad multivariado, generalmente se procede analizando la normalidad de las marginales del vector $\underline{\boldsymbol{x}}_{p\times 1}$ y analizando la normalidad bivariada de pares de componentes de dicho vector.

En la práctica, no es frecuente encontrar conjuntos de datos normales en bajas dimensiones (ie, $p = 1$ o $p = 2$) y que no lo sean en altas dimensiones, pero hay que tener en cuenta que **la normalidad univariada no implica la normalidad multivariada**, ver un ejemplo en el EJERCICIO 4.8 de Johnson, un caso de normalidad univariadas que no implica Normalidad-bi-variada.

Evaluación a nivel marginal (ie. Normalidad Univariada)

Existen varios enfoques para evaluar la normalidad Uni-variada, entre ellos están los siguientes.

1. Gráficos como histogramas, cajas de bigote, etc.
2. Gráficos *qq*-plot y *NPP*.
3. Prueba de Normalidad basada en el coeficiente de correlación de los puntos del *qq*-plot.
4. Análisis de las combinaciones lineales de las variables del vector.
5. Pruebas formales de Normalidad.

A continuación se explican de manera breve cada uno de estos procedimientos.

1. Gráficos como histogramas, cajas de bigote, etc.

Generalmente se utilizan los histogramas o cajas de bigotes cuando la muestra es de tamaño moderado o grande y los diagramas de puntos en el caso de n-pequeño, para detectar alejamientos de la simetría de los datos, una cola parece ser mayor que otra.

2. Gráficos qq-plot.

Uso del gráficos cuantil-cuantil o qq-plot. Estos Son gráficos especiales que pueden ser usados para evaluar la normalidad de cada variable.

En ellos se grafican los cuantiles muestrales contra los cuantiles que se esperaría observar si las observaciones realmente provienen de una distribución normal.

Los pasos para construir un $Q-Q$-plot son:

a) Ordene las observaciones originales de menor a mayor.

Sean $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$, las observaciones ordenadas.

Las probabilidades correspondientes a ellos son:

$$\left(1-\frac{1}{2}\right)/n, \left(2-\frac{1}{2}\right)/n, \cdots, \left(n-\frac{1}{2}\right)/n$$

Lo anterior quiere decir, que la proporción i/n-de la muestra que está a la izquierda de $X_{(i)}$ se aproxima por: $\frac{i-\frac{1}{2}}{n}$.

b) Se calculan los cuantiles de la normal estándar: $q_{(1)}, q_{(2)}, \cdots, q_{(n)}$, correspondientes a las probabilidades anteriores.

c) Grafique los pares de observaciones:

$$[q_{(1)}, X_{(1)}]\ ,\ [q_{(2)}, X_{(2)}]\ ,\ \cdots\ ,\ [q_{(n)}, X_{(n)}].$$

Si los datos proceden de una distribución normal, se espera que el gráfico anterior sea aproximadamente una línea recta.

Lo anterior significa que los pares $[q_{(i)}, X_{(i)}]$
estarán aproximadamente relacionados de forma lineal, ya que
$\sigma q_{(j)} + \mu$ estará muy cerca del cuantil muestral esperado.

3. Gráficos *NPP*.

Usar el Normal-probability-plot (*NPP*), para el cual se grafican las parejas, $[m_{(i)}\ ,\ x_{(i)}]$, donde $m_{(i)} = E[Z_{(i)}]$-es el valor esperado del i-ésimo estadístico de orden en una muestra de tamaño n de una normal estándar.

Este gráfico debe dar similar al *qq*-plot, ie. una línea recta.

Estos gráficos *qq*-plot y NPP no son muy claro, a menos que los tamaños de muestra sean moderadamente grandes, (ie. por ejemplo $n \geq 20$), ya que pueden mostrar observaciones muy alejadas de una tendencia lineal, aún cuando se sabe que los datos provienen de una distribución normal.

Ejemplo-1

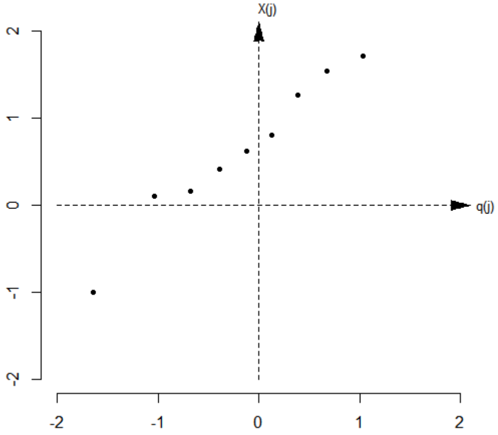
Considere una muestra de tamaño $n = 10$ observaciones, las cuales fueron ordenadas de menor a mayor como se muestra en la siguiente tabla.

Por ejemplo, para el cálculo del cuantil de la $N(0,1)$, para una probabilidad de 0.65 busca el cuantil que satisface:
$P[Z < q_{(7)}] = 0.65$,

de donde se tiene que dicho cuantil es: $q(7) = 0.385$, puesto que:

$$P[Z < 0.385] = \int_{-\infty}^{0.385} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = 0.65.$$

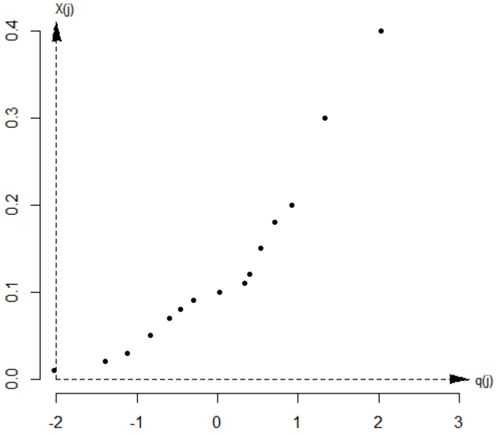
Observaciones Ordenadas $x_{(j)}$	Probabilidades $\left(j - \frac{1}{2}\right)/n$	Quantiles de la $N(0,1)$ $q_{(j)}$
-1.00	0.05	-1.645
-0.10	0.15	-1.036
0.16	0.25	-0.674
0.41	0.35	-0.385
0.62	0.45	-0.125
0.80	0.55	0.125
1.26	0.65	0.385
1.54	0.75	0.674
1.71	0.85	1.036
2.30	0.95	1.645



La construcción del *qq*-plot se basa en el diagrama de dispersión de los puntos $(q_{(j)}, x_{(j)}), \;\; j = 1, 2, \cdots, 10$, los cuales caen muy cerca de una línea recta, lo que conduce a no rechazar la hipótesis de que estos datos son de una distribución normal.

Ejemplo-2 El departamento de control de calidad de una empresa que produce hornos micro-ondas requiere monitorear la cantidad de radiación emitida por dichos micro-ondas cuando tienen la tapa cerrada. Aleatoriamente se eligieron $n = 42$ hornos y se observó dicha cantidad de radiación emitida por ellos con la tapa cerrada.

Horno	Radiación	Horno	Radiación	Horno	Radiación	Horno	Radiación
1	0.15	12	0.02	23	0.03	34	0.30
2	0.09	13	0.01	24	0.05	35	0.02
3	0.18	14	0.10	25	0.15	36	0.20
4	0.10	15	0.10	26	0.10	37	0.20
5	0.05	16	0.10	27	0.15	38	0.30
6	0.12	17	0.02	28	0.09	39	0.30
7	0.08	18	0.10	29	0.08	40	0.40
8	0.05	19	0.01	30	0.18	41	0.30
9	0.08	20	0.40	31	0.10	42	0.05
10	0.10	21	0.10	32	0.20		
11	0.07	22	0.05	33	0.11		



La apariencia del gráfico indica que los datos no parecen provenir de una distribución normal. Los puntos señalados con un círculo son observaciones atípicas, pues están muy lejos del resto de los datos.

Observación: Para esta muestra de datos de radiación, varias observaciones son iguales (observaciones empatadas). Cuando esto ocurre, a las observaciones con valores iguales se les asigna un mismo cuantil, el cual se obtiene usando el promedio de los cuantiles que ellas hubieran tenido si hubieran sido ligeramente distintas.

4. Prueba de Normalidad basada en el coeficiente de correlación de los puntos del *qq*-plot.

La linealidad de un gráfico *qq*-plot puede ser medida calculando el coeficiente de correlación para los puntos de dicho gráfico, el cual esta dado por:

$$r_Q = \frac{\sum_{j=1}^{n}(x_{(j)} - \overline{x})(q_{(j)} - \overline{q})}{\sqrt{\sum_{j=1}^{n}(x_{(j)} - \overline{x})^2}\sqrt{\sum_{j=1}^{n}(q_{(j)} - \overline{q})^2}}$$

Basados en este coeficiente de correlación, se puede construir una prueba potente de normalidad (ver. Filliben, 1975; Looney y Gulledge, 1985; Shapiro y Wilk, 1965). Formalmente, se rechaza la hipótesis de normalidad a un nivel de significancia de α si $r_Q < r_Q(\alpha, n)$ donde los valores críticos $r_Q(\alpha, n)$ se encuentran en la siguiente tabla:

Valores críticos para el coeficiente de correlación del gráfico *qq*-plot para probar normalidad.

Tamaño de muestra n	Nivel de Significancia α		
	0.01	0.05	0.10
5	0.8299	0.8788	0.9032
10	0.8801	0.9198	0.9351
15	0.9126	0.9389	0.9503
20	0.9269	0.9508	0.9604
25	0.9410	0.9591	0.9665
30	0.9479	0.9652	0.9715
35	0.9538	0.9682	0.9740
40	0.9599	0.9726	0.9771
45	0.9632	0.9749	0.9792
50	0.9671	0.9768	0.9809
55	0.9695	0.9787	0.9822
60	0.9720	0.9801	0.9836
75	0.9771	0.9838	0.9866
100	0.9822	0.9873	0.9895
150	0.9879	0.9913	0.9928
200	0.9905	0.9931	0.9942
300	0.9935	0.9953	0.9960

Ejemplo-3.

Para el primer ejemplo donde $n = 10$, el cálculo del coeficiente de correlación entre los puntos $(q_{(j)}, X_{(j)})$, $j = 1, 2, \cdots, 10$, del gráfico qq-plot es:

$$r_Q = \frac{\sum_{j=1}^{10}(x_{(j)} - \overline{x})(q_{(j)} - \overline{q})}{\sqrt{\sum_{j=1}^{10}(x_{(j)} - \overline{x})^2}\sqrt{\sum_{j=1}^{10}(q_{(j)} - \overline{q})^2}} = \frac{8.584}{\sqrt{8.472}\sqrt{8.795}} = 0.994,$$

de donde, para un nivel de significancia $\alpha = 0.10$, el valor crítico de la tabla es: $r_Q(0.10, 10) = 0.9351$, luego como $r_Q = 0.994 > 0.9351 = r_Q(0.10, 10)$, entonces no rechazamos la hipótesis de normalidad.

Observación: Para muestras grandes, las pruebas basadas en r_Q y la prueba de Shapiro Wilk (una prueba potente de normalidad) son aproximadamente las mismas.

5. Analisis de las combinaciones lineales de las variables de $\underline{\boldsymbol{x}}$

Considere los valores propios de $\boldsymbol{S}$, $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_p$ y sus correspondientes vectores propios $\underline{\hat{\boldsymbol{e}}}_1, \underline{\hat{\boldsymbol{e}}}_2, \ldots, \underline{\hat{\boldsymbol{e}}}_p$. Se sugiere verificar normalidad para las combinaciones lineales dadas por:

$$\underline{\hat{\boldsymbol{e}}}_1^t \underline{\boldsymbol{x}}_j \quad \text{y} \quad \underline{\hat{\boldsymbol{e}}}_p^t \underline{\boldsymbol{x}}_j$$

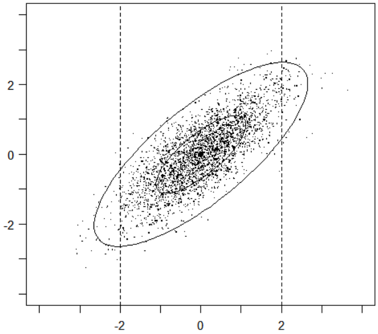
donde $\underline{\hat{\boldsymbol{e}}}_1$ y $\underline{\hat{\boldsymbol{e}}}_p$ son los vectores propios correspondientes al mayor y menor valor propio de $\boldsymbol{S}$, respectivamente.

6. Pruebas Formales.

Además de todas las ayudas anteriores, se pueden hacer pruebas formales de normalidad para el caso univariado, como lo son: Kolmogorov-Smirnov, Shapiro-Wilks,Anderson-Darling, Cramer-Von-Mises, etc.

Evaluación de la Normalidad Bi-variada

Si las observaciones fueran generadas por un distribución normal multivariada ($p > 2$), todas las distribuciones bivariadas ($p = 2$) deberían ser normales y los contornos de densidad constante deberían se elipses. Observe el siguiente diagrama de dispersión generado por una muestra simulada de una normal bivariada.



Además, por resultado mencionado anteriormente, el conjunto de puntos bivariados $\underline{\boldsymbol{x}}$ tales que:

$$(\underline{X} - \underline{\mu})^t \boldsymbol{\Sigma}^{-1} (\underline{X} - \underline{\mu}) \leq \chi^2_\alpha(2)$$

tiene una probabilidad $1 - \alpha$, ie.

$$P\left[(\underline{X} - \underline{\mu})^t \boldsymbol{\Sigma}^{-1} (\underline{X} - \underline{\mu}) \leq \chi^2_\alpha(2)\right] = 1 - \alpha.$$

Por ejemplo, si $\alpha = 0.5$, para muestras grandes se esperaría que alrededor del 50% de las observaciones caigan dentro de la elipse dada por:

$$(\underline{X} - \underline{\overline{\boldsymbol{x}}})^t S^{-1} (\underline{X} - \underline{\overline{\boldsymbol{x}}}) \leq \chi^2_{0.5}(2) = 1.39$$

Si no se cumple esto, entonces la normalidad bi-variada es sospechosa.

Ejemplo-4 Considere los pares de datos para las variables X_1 =ventas y X_2 =ganancias para las 10 mayores corporaciones industriales de E.U.

Compañía	X_1-Ventas (Billones)	X_2-Ganancias (Billones)	X_3-Activo (Billones)
Citigroup	108.28	17.05	1484.10
General Electric	152.36	16.59	750.33
American Intl Group	95.04	10.91	766.42
Bank of America	65.45	14.14	1110.46
HSBC Group	62.97	9.52	1031.29
ExxonMobil	263.99	25.33	195.26
Royal Dutch/Shell	265.19	18.54	193.83
BP	285.06	15.73	191.11
ING Group	92.01	8.10	1175.16
Toyota Motor	165.68	11.13	211.15

Para estos datos se tiene que:

$$\underline{\overline{\boldsymbol{x}}} = \begin{bmatrix} 155.60 \\ 14.70 \end{bmatrix} , \qquad S = \begin{bmatrix} 7476.45 & 303.62 \\ 303.62 & 26.19 \end{bmatrix}$$

$$S^{-1} = \frac{1}{103623.12} \begin{bmatrix} 26.19 & -303.62 \\ -303.62 & 7476.45 \end{bmatrix}$$

Para $\alpha = 0.5$, a partir de la distribución chi-cuadrado con $p = 2$-grados de libertad, se obtiene que: $\chi^2_{0.5}(2) = 1.39$, de donde, cualquier observación $\underline{\boldsymbol{x}} = (x1, x2)$ que cumpla la siguiente desigualdad:

$$\begin{bmatrix} x_1 - 155.60 & x_2 - 14.70 \end{bmatrix} \left(\begin{bmatrix} 0.000253 & -0.002930 \\ -0.002930 & 0.072148 \end{bmatrix} \times 10^{-5} \right) \begin{bmatrix} x_1 - 155.60 \\ x_2 - 14.70 \end{bmatrix} \leq 1.39,$$

debe estar sobre o dentro del contorno estimado del 50% de probabilidad. De lo contrario la observación estaría por fuera.

Para las 10 observaciones observadas se tiene que sus distancias generalizadas son:

$1.61, \mathbf{0.30}, \mathbf{0.62}, 1.79, \mathbf{1.30}, 4.38, 1.64, 3.53, 1.71$ y $\mathbf{1.16}$.

Si los datos proceden de una distribución normal, se esperaría que aproximadamente el 50% de las observaciones caiga dentro o sobre el contorno estimado anterior, o dicho de otro modo, el 50% de las distancias calculadas deberían ser menores o iguales que 1.39.

Se observa que en total 4 de estas 10 distancias son menores que 1.39, lo que implica que la proporción estimada de observaciones que cumplen la desigualdad es del 40%.

La diferencia entre de esta proporción con el 50% (ie. un 10% de diferencia) proporciona una evidencia para rechazar la normalidad bivariada en estos datos. **Sin embargo, la muestra es muy pequeña para permitir obtener esta conclusión.**

Otro Método:

El procedimiento anterior es útil, pero bastante burdo. Un método más formal para evaluar la normalidad conjunta está basado en las distancias cuadráticas generalizadas, dadas por:

$$d_j^2 = (\underline{X}_j - \underline{\overline{\boldsymbol{x}}})^t S^{-1} (\underline{X}_j - \underline{\overline{\boldsymbol{x}}}) \text{ , para } j = 1, 2, \ldots, n,$$

donde, $\underline{X}_j$-son las observaciones muestrales.

El siguiente procedimiento, no está limitado al caso bi-variado, por lo que puede ser usado par $p \geq 2$.

Cuando la población es normal multivariada y cuando tanto n como $n - p$ son grandes, por ejemplo, (≥ 25 *o* 30), cada una de las distancias al cuadrado d_j^2-anterior, deberían comportarse como una variable aleatoria con distribución χ^2.

Aunque, estas distancias no son independientes o distribuidas chi-cuadrados exactamente, es útil graficarlas como si lo fueran.

El gráfico resultante es llamado Gráfico Chi-Cuadrado, y se construye de la siguiente manera:

- Se Ordenan las distancias de menor a mayor: $d^2_{(1)} \leq d^2_{(2)} \leq \ldots d^2_{(n)}$
- Se grafican los pares: $\left(q_{c,p}\left(\frac{j-\frac{1}{2}}{n}\right)\ ,\ d^2_j\right)\ ,\ j = 1, 2, \ldots, n$

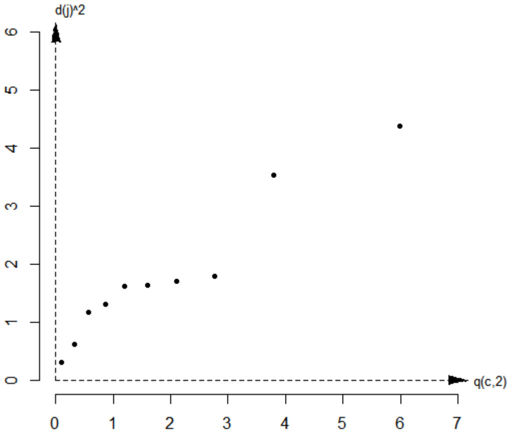
  donde, $q_{c,p}\left(\frac{j-\frac{1}{2}}{n}\right)$-es el cuantil $100\left(j - \frac{1}{2}\right)/n$ de la distribución chi-cuadrado con p-grados de libertad, ie.

$$P\left[\chi^2_{(p)} \leq q_{c,p}\left(\frac{j-\frac{1}{2}}{n}\right)\right] = \frac{j-\frac{1}{2}}{n}$$

Bajo normalidad, el gráfico debería mostrar un patrón lineal a través del origen y con pendiente 1. Un patrón sistemáticamente curvo sugiere falta de normalidad multivariada.

Ejemplo-5 La figura tiene las distancias ordenadas y los percentiles chi-cuadrado al igual que el gráfico chi- cuadrado, para el ejemplo anterior.

Observaciones j	Distancias Ordenadas $d^2_{(j)}$	Quantiles χ^2_2 de $q_{c,2}\left(\frac{j-\frac{1}{2}}{10}\right)$
1	0.30	0.10
2	0.62	0.33
3	1.16	0.58
4	1.30	0.86
5	1.61	1.20
6	1.64	1.60
7	1.71	2.10
8	1.79	2.77
9	3.53	3.79
10	4.38	5.99

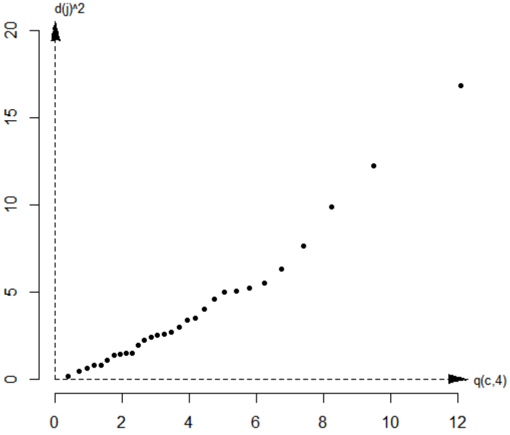


Se observa que los puntos no caen en una línea recta de pendiente 1. Por lo que no se apoya la normalidad bi-variada en estos datos. Aunque hay que tener en cuenta que n es pequeño.

Ejemplo-6 La figura tiene las distancias ordenadas y los percentiles chi-cuadrado al igual que el gráfico chi- cuadrado, para una muestra 4-variada.

	x_1	x_2	x_3	x_4	d_2	$q(c,4)$
1	1889	1651	1561	1778	0.60	0.39
2	2403	2048	2087	2197	5.48	0.71
3	2119	1700	1815	2222	7.62	0.95
4	1645	1627	1110	1533	5.21	1.17
5	1976	1916	1614	1883	1.40	1.37
6	1712	1712	1439	1546	2.22	1.56
7	1943	1685	1271	1671	4.99	1.74
8	2104	1820	1717	1874	1.49	1.92
9	2983	2794	2412	2581	12.26	2.10
10	1745	1600	1384	1508	0.77	2.29
11	1710	1591	1518	1667	1.93	2.47
12	2046	1907	1627	1898	0.46	2.66
13	1840	1841	1595	1741	2.70	2.85
14	1867	1685	1493	1678	0.13	3.05
15	1859	1649	1389	1714	1.08	3.25

	x_1	x_2	x_3	x_4	d_2
16	1954	2149	1180	1281	16.85
17	1325	1170	1002	1176	3.50
18	1419	1371	1252	1308	3.99
19	1828	1634	1602	1755	1.36
20	1725	1594	1313	1646	1.46
21	2276	2189	1547	2111	9.90
22	1899	1614	1422	1477	5.06
23	1633	1513	1290	1516	0.80
24	2061	1867	1646	2037	2.54
25	1856	1493	1356	1533	4.58
26	1727	1412	1238	1469	3.40
27	2168	1896	1701	1834	2.38
28	1655	1675	1414	1597	3.00
29	2326	2301	2065	2234	6.28
30	1490	1382	1214	1284	2.58



Se observa que los puntos no caen en una línea recta de pendiente 1. Por lo que no se apoya la normalidad bi-variada en estos datos. Aunque hay que tener en cuenta que n es pequeño.

NOTA: Consultar pruebas multivariadas de asimetría y kurtosis.

Detección de Observaciones Atípicas

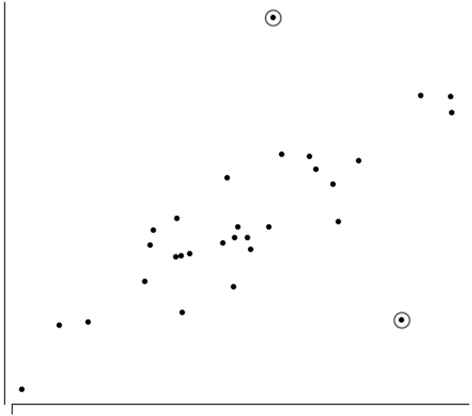
La mayoría de los conjuntos de datos contienen unas pocas observaciones inusuales que no parecen pertenecer al patrón de variabilidad seguido por las otras observaciones.

Estas observaciones son denominadas Observaciones Atípicas y antes de proceder a identificarlas se debe enfatizar que no todas las observaciones atípicas son números equivocados.

Ellas pueden formar parte del grupo y conducir a comprender mejor el fenómeno que se estudia.

La detección de observaciones atípicas puede ser mejor realizada visualmente, es decir por medio de gráficos.

Para el caso de una variable, se pueden usar gráficos de puntos para muestras pequeñas y gráficos de cajas para muestras grandes. Para el caso de dos variables analicemos el siguiente gráfico.



Para el Caso Bivariado el diagrama de dispersión proporciona la información visual requerida para detectar datos atípicos. Sin embargo, en altas dimensiones, los datos atípicos pueden no ser detectados por gráficos uni-variados o aún diagramas de dispersión.

En estos casos se recomienda usar algunos tipos de gráficos para datos multivariados como los son: las curvas de Andrews, las gráficas de caras de Chernoff y gráficos de estrellas, etc (Consultar sección 1.4 del libro de Johnso). Estos gráficos son muy potentes para detectar casos atípicos multivariados.

Además, en altas dimensiones un valor grande de las distancias:

$$d_j^2 = (\underline{X}_j - \underline{\overline{\boldsymbol{x}}})^t S^{-1} (\underline{X}_j - \underline{\overline{\boldsymbol{x}}}) \ , \ \text{para} \ \ j = 1, 2, \ldots, n,$$

sugieren que la observación $\underline{X}_j$ es inusual, aunque no la hallamos visualizado gráficamente.

Resumen: Los pasos para detectar observaciones extremas o Outliers en datos multivariados son:

a) Hacer gráficos de puntos para cada variable.

b) Hacer gráficos de dispersión para cada par de variables.

c) Calcular los valores estandarizados de cada variable dados por:

$$z_{jk} = \frac{(x_{jk} - \overline{x}_k)}{\sqrt{s_{kk}}} \text{ , para } j = 1, 2, \ldots, n \text{ y } k = 1, 2, \ldots, p.$$

Examinar aquellos valores estandarizados muy grandes o muy pequeños.

d) Calcular las distancias cuadradas generalizadas:

$$d_j^2 = (\underline{X}_j - \underline{\overline{\mathbf{x}}})^t S^{-1} (\underline{X}_j - \underline{\overline{\mathbf{x}}}) \text{ , para } j = 1, 2, \ldots, n.$$

Examinar aquellas distancias de valores muy grandes.
En el gráfico chi cuadrado, son aquellos puntos más alejados del origen.

Transformaciones para acercar a la normalidad multivariada

Cuando la normalidad no es un supuesto viable, en algunos casos se pueden hacer transformaciones de los datos para acercarlos a la normalidad.

Las transformaciones son solamente reexpresiones de los datos en diferentes unidades. Por ejemplo, cuando un histograma de observaciones positivas muestra una gran cola derecha, una transformación de ellos tomando el logaritmo o la raíz cuadrada generalmente mejora la simetría con respecto a la media y aproxima la distribución a la normalidad.

Los tipo de transformaciones a realizar, pueden ser sugeridos por las características de los mismos datos o por consideraciones teóricas. En el caso de transformaciones sugeridas por los mismos datos, se tiene por ejemplo que, los datos de conteos pueden ser más normales si se les toma la raíz cuadrada. Similarmente, para datos de proporciones la transformación logit y para datos de correlación la transformación de Fisher.

Resumen:

Escala Original	Escala Transformada
Conteos, y	$\sqrt{y}$
Proporciones $\hat{p}$	$Logit(\hat{p}) = \frac{1}{2}Log\left(\frac{\hat{p}}{1-\hat{p}}\right)$
Correlaciones, r	Fisher: $z(r) = \frac{1}{2}Log\left(\frac{1+r}{1-r}\right)$

En el caso de consideraciones teóricas,
una familia de transformaciones para este propósito es la familia de transformaciones de potencias. Existe un método analítico conveniente para escoger una transformación de potencia dentro de dicha familia.

Familia de transformaciones de potencia de Box y Cox (1964)

$$x^{(\lambda)} = \begin{cases} \frac{x^{\lambda}-1}{\lambda} \ , & \text{para} \quad \lambda \neq 0 \\ \mathrm{Ln}(x) \ , & \text{para} \quad \lambda = 0 \end{cases}$$

La cual es continua en λ para $x > 0$.

Dadas las observaciones $x_1, x_2, \ldots, x_n$, la solución de Box y Cox para escoger la transformación λ adecuada (ie. el valor de λ adecuado), es aquella que maximiza la expresión:

$$l(\lambda) = -\frac{n}{2}\mathrm{Ln}\left[\frac{1}{n}\sum_{j=1}^{n}\left(x_j^{(\lambda)} - \overline{x^{(\lambda)}}\right)^2\right] + (\lambda - 1)\sum_{j=1}^{n}\mathrm{Ln}(x_j) \ , \quad \text{donde,}$$

$$\overline{x^{(\lambda)}} = \frac{1}{n}\sum_{j=1}^{n} x_j^{(\lambda)} = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{x_j^{(\lambda)} - 1}{\lambda}\right),$$

es la media aritmética de las observaciones transformadas.

El proceso de maximización de $l(\lambda)$ es fácil de realizar por medio de un computador, seleccionando muchos valores diferentes para λ y calculando el respectivo valor de $l(\lambda)$.

Es útil hacer un gráfico de $l(\lambda)$ versus λ para estudiar el comportamiento en el valor del máximo $\hat{\lambda}$.

Algunos autores, recomiendan un procedimiento equivalente para encontrar λ, creando una nueva variable definida por:

$$y_j^{(\lambda)} = \frac{x_j^{\lambda} - 1}{\lambda \left[\left(\prod_{i=1}^{n} x_i \right)^{1/n} \right]^{\lambda - 1}} \ , \quad j = 1, 2, \ldots, n$$

y calculando su varianza muestral. El mínimo de la varianza ocurre en el mismo valor λ que maximiza $l(\lambda)$.

Ejemplo-7 Para un ejemplo visto anteriormente de $n = 42$ datos de la radiación de hornos de micro-ondas con la tapa cerrada, el gráfico $qq - plot$ indica que las observaciones se desvían de lo que esperaríamos si fueran normalmente distribuidas.

Puesto que todas las observaciones son positivas, se puede utilizar una transformación de potencia de los datos con la esperanza de acercarlos a la normalidad.

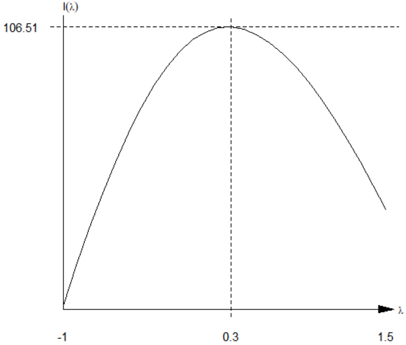
Los pares $\Big(\lambda\ ,\ l(\lambda)\Big)$, para este proceso de búsqueda se encuentran en la siguiente tabla, al igual que el gráfico de $l(\lambda)$ contra λ, donde observamos que el máximo se alcanza en $\hat{\lambda} = 0.28$, pero por conveniencia elegimos a $\hat{\lambda} = 0.3$.

	λ	$I(\lambda)$
1	-1.00	70.52
2	-0.90	75.62
3	-0.80	80.46
4	-0.70	84.94
5	-0.60	89.06
6	-0.50	92.79
7	-0.40	96.10
8	-0.30	98.97
9	-0.20	101.39

	λ	$I(\lambda)$
10	-0.10	103.35
11	0.00	104.83
12	0.10	105.84
13	0.20	106.39
14	0.30	106.51
15	0.40	106.20
16	0.50	105.50
17	0.60	104.43
18	0.70	103.03

	λ	$I(\lambda)$
19	0.80	101.33
20	0.90	99.34
21	1.00	97.10
22	1.10	94.64
23	1.20	91.96
24	1.30	89.10
25	1.40	86.07
26	1.50	82.88

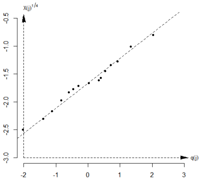
Gráfica de: $l(\lambda)$ $v.s$ λ

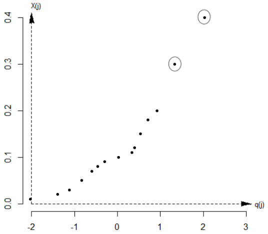


Luego, los datos son transformados como:

$$x_j^{(\lambda)} = \frac{x_j^{(0.3)} - 1}{0.3} \ , \ \text{para } j = 1, 2, \ldots, 42.$$

Para verificar si los datos transformados están más cerca a la normal, se presenta su gráfico $qq-plot$ para datos transformados y no-transformados.





Los pares de cuantiles caen muy cerca de una recta, lo que permite concluir que los datos: $x_j^{(\lambda)}$ son aproximadamente normal.

Transformaciones para el Caso Multivariado

Para las observaciones multivariadas se debe seleccionar una transformación para cada variable.

Sea $\lambda_1, \lambda_2, \ldots, \lambda_p$ las transformaciones de potencia para las p variables consideradas.

Las transformaciones pueden ser obtenidas individualmente para cada una de las variables siguiendo el procedimiento anterior.

La j-ésima observación transformada esta dada por:

$$\underline{x_j}^{(\hat{\lambda})} = \begin{bmatrix} \frac{x_{j1}^{(\hat{\lambda}_1)}-1}{\hat{\lambda}_1} \\ \\ \frac{x_{j2}^{(\hat{\lambda}_2)}-1}{\hat{\lambda}_2} \\ \vdots \\ \frac{x_{jp}^{(\hat{\lambda}_p)}-1}{\hat{\lambda}_p} \end{bmatrix}$$

donde $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_p$ son los valores que individualmente maximizan a $l(\lambda_k)$, para $k = 1, 2 \ldots, p$.

El anterior procedimiento es equivalente a hacer cada distribución marginal aproximadamente normal. Aunque, la normalidad marginal de cada componente no es suficiente para garantizar que todas la distribución conjunta sea normal multivariada, frecuentemente esta condición es suficiente.

Muestreo en la distribución normal p-variada

Sea una muestra aleatoria normal p-variada de tamaño n, $(\underline{\boldsymbol{x}}_1, \underline{\boldsymbol{x}}_2, \ldots, \underline{\boldsymbol{x}}_n)$ con vector de medias $\underline{\boldsymbol{\mu}}$ y matriz de var-cov $\boldsymbol{\Sigma}$, es decir:

$$\mathbf{X}_{n \times p} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix} = \begin{bmatrix} \underline{\boldsymbol{x}}_1 \\ \underline{\boldsymbol{x}}_2 \\ \vdots \\ \underline{\boldsymbol{x}}_n \end{bmatrix}$$

Se tienen n-vectores $\underline{\boldsymbol{x}}_i$ para $i = 1, 2, \cdots, n$ en $\mathbb{R}^p$ con media dada por:

$E[\underline{\boldsymbol{x}}_i] = \underline{\boldsymbol{\mu}}_{p \times 1}$ y matriz de var-cov dada por: $Var[\underline{\boldsymbol{x}}_i] = \boldsymbol{\Sigma}_{p \times p}$.

Estimadores de Máx-Ver de una Normal-Multivariada

Muchos procesos estadísticos emplean los valores de los parámetros poblacionales que mejor expliquen los datos observados, en donde **mejor**, se refiere a elegir valores de $\underline{\boldsymbol{\mu}}$ y $\boldsymbol{\Sigma}$ que maximicen la f.d.p conjunta evaluada en los datos observados, ie. maximizan la probabilidad de los datos.

A la técnica de hallar estos parámetros poblacionales que maximizan la f.d.p conjunta para datos observados fijos, se le conoce con el nombre de: Estimación de Máxima Verosimilitud y a los parámetros estimados mediante esta técnica se les llaman Estimadores de Máxima Verosimilitud (MLE) o de máxima probabilidad.

A continuación se hallarán los MLE de los parámetros poblacionales $\underline{\boldsymbol{\mu}}$ y $\boldsymbol{\Sigma}$ de una población Normal-Multivariada, para un conjunto de datos u observaciones fijas: $\underline{\boldsymbol{x}}_1, \underline{\boldsymbol{x}}_2, \ldots, \underline{\boldsymbol{x}}_n$ p-variadas.

Función de densidad de probabilidad conjunta de una N-Multivariada

Como las $\underline{\boldsymbol{x}}_i$ para $i = 1, 2, \ldots, n$, son mutuamente independientes y cada uno tiene una distribución normal p-variada, es decir $\underline{\boldsymbol{x}}_i \sim N_p(\underline{\boldsymbol{\mu}}\ ,\ \boldsymbol{\Sigma})$, entonces la f.d.p conjunta de las n-observaciones p-variadas es el producto de las densidades normales (p-variadas) marginales, es decir:

$$f(\ \underline{\boldsymbol{x}_1}, \underline{\boldsymbol{x}_2}, \cdots, \underline{\boldsymbol{x}_n}\ ;\ \underline{\boldsymbol{\mu}}\ ,\ \boldsymbol{\Sigma}\ ) = \prod_{j=1}^{n} f(\underline{\boldsymbol{x}}_j)$$

$$= \prod_{j=1}^{n} \left\{ \frac{1}{(2\pi)^{p/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \mathrm{e}^{-\frac{1}{2}(\underline{\boldsymbol{x}}_j - \underline{\boldsymbol{\mu}})' \boldsymbol{\Sigma}^{-1} (\underline{\boldsymbol{x}}_j - \underline{\boldsymbol{\mu}})} \right\}$$

$$= \frac{1}{(2\pi)^{np/2}} \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} \mathrm{e}^{-\frac{1}{2}\sum_{j=1}^{n}(\underline{\boldsymbol{x}}_j - \underline{\boldsymbol{\mu}})' \boldsymbol{\Sigma}^{-1} (\underline{\boldsymbol{x}}_j - \underline{\boldsymbol{\mu}})}$$

Función de verosimilitud de una Normal Multivariada

Para valores conocidos de las observaciones:
$(\underline{\boldsymbol{x}}_1 = \underline{x}_1, \ldots, \ldots, \underline{\boldsymbol{x}}_n = \underline{x}_n)$, la f.d.p conjunta anterior, se convierte en una función de los parámetros $\underline{\boldsymbol{\mu}}$ y $\boldsymbol{\Sigma}$, la cual se llama: Función de Verosimilitud y se denota por $L(\underline{\boldsymbol{\mu}}\ ,\ \boldsymbol{\Sigma})$, ie:

$$L(\ \underline{\boldsymbol{\mu}}\ ,\ \boldsymbol{\Sigma}\ ) = f(\ \underline{\boldsymbol{\mu}}\ ,\ \boldsymbol{\Sigma} \mid \underline{x}_1, \underline{x}_1, \ldots, \underline{x}_n\ )$$

$$= \frac{1}{(2\pi)^{np/2}} \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} e^{-\frac{1}{2}\sum_{j=1}^{n}(\underline{x}_j - \underline{\boldsymbol{\mu}})' \boldsymbol{\Sigma}^{-1}(\underline{x}_j - \underline{\boldsymbol{\mu}})}$$

Para simplificar el exponente de la función de verosimilitud anterior, veámos algunas propiedades de la traza de una matriz.

Teorema: Sea $\boldsymbol{A}_{p\times p}$, una matriz simétrica y $\underline{\boldsymbol{x}}_{p\times 1}$ un vector p-variado, entonces:

- $\underline{\boldsymbol{x}}'\boldsymbol{A}\underline{\boldsymbol{x}} = \operatorname{tr}\left(\underline{\boldsymbol{x}}'\boldsymbol{A}\underline{\boldsymbol{x}}\right) = \operatorname{tr}\left(\boldsymbol{A}\underline{\boldsymbol{x}}\underline{\boldsymbol{x}}'\right)$ *tr* :-función traza.
- $\operatorname{tr}(\boldsymbol{A}) = \sum_{i=1}^{p} \lambda_i$, con λ_i: valores propios de $\boldsymbol{A}$

Ahora se utilizará este resultado para simplificar el exponente mensionado anteriormente.

Simplificación del Exponente de la función de verosimilitud

$$
\begin{aligned}
(\underline{x}_j - \underline{\boldsymbol{\mu}})' \boldsymbol{\Sigma}^{-1} (\underline{x}_j - \underline{\boldsymbol{\mu}}) &= \operatorname{tr}\left[(\underline{x}_j - \underline{\boldsymbol{\mu}})' \boldsymbol{\Sigma}^{-1} (\underline{x}_j - \underline{\boldsymbol{\mu}})\right] \\
&= \operatorname{tr}\left[\boldsymbol{\Sigma}^{-1} (\underline{x}_j - \underline{\boldsymbol{\mu}})(\underline{x}_j - \underline{\boldsymbol{\mu}})'\right]
\end{aligned}
$$

$$
\begin{aligned}
\text{Luego,} \sum_{j=1}^{n} (\underline{x}_j - \underline{\boldsymbol{\mu}})' \boldsymbol{\Sigma}^{-1} (\underline{x}_j - \underline{\boldsymbol{\mu}}) &= \sum_{j=1}^{n} \operatorname{tr}\left[\boldsymbol{\Sigma}^{-1} (\underline{x}_j - \underline{\boldsymbol{\mu}})(\underline{x}_j - \underline{\boldsymbol{\mu}})'\right] \\
&= \operatorname{tr}\left[\boldsymbol{\Sigma}^{-1} \underbrace{\left(\sum_{j=1}^{n} (\underline{x}_j - \underline{\boldsymbol{\mu}})(\underline{x}_j - \underline{\boldsymbol{\mu}})'\right)}\right]
\end{aligned}
$$

Ahora, se suma y se resta: $\underline{\bar{\boldsymbol{x}}} = \frac{1}{n}\boldsymbol{X}^T\mathbf{1}$ en cada término:
$(\underline{x}_j - \underline{\boldsymbol{\mu}})$ de la suma: $\sum_{j=1}^{n} (\underline{x}_j - \underline{\boldsymbol{\mu}})(\underline{x}_j - \underline{\boldsymbol{\mu}})'$, con lo cual se obtiene:

$$\sum_{j=1}^{n}(\underline{x}_j - \underline{\boldsymbol{\mu}})'\boldsymbol{\Sigma}^{-1}(\underline{x}_j - \underline{\boldsymbol{\mu}}) = \operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\underbrace{\left(\sum_{j=1}^{n}(\underline{x}_j - \underline{\boldsymbol{\mu}})(\underline{x}_j - \underline{\boldsymbol{\mu}})'\right)}\right]$$

$$= \operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\underbrace{\left(\sum_{j=1}^{n}(\underline{x}_j - \underline{\overline{\boldsymbol{x}}} + \underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})(\underline{\boldsymbol{x}}_j - \underline{\overline{\boldsymbol{x}}} + \underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})'\right)}\right]$$

$$= \operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\underline{x}_j - \underline{\overline{\boldsymbol{x}}})(\underline{x}_j - \underline{\overline{\boldsymbol{x}}})' + \sum_{j=1}^{n}(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})'\right.\right.$$

$$\left.\left. + \underbrace{\sum_{j=1}^{n}(\underline{x}_j - \underline{\overline{\boldsymbol{x}}})(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})'}_{=0} + \underbrace{\sum_{j=1}^{n}(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})(\underline{x}_j - \underline{\overline{\boldsymbol{x}}})'}_{=0}\right)\right]$$

Expresión final del exponente de la f.d.p.c

Pero los últimos dos términos sub-rayados en la expresión anterior son iguales a cero, de donde se obtiene que el exponente final simplificado de la función de verosimilitud es:

$$\sum_{j=1}^{n}(\underline{x}_j - \underline{\boldsymbol{\mu}})' \boldsymbol{\Sigma}^{-1}(\underline{x}_j - \underline{\boldsymbol{\mu}})$$

$$= \operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\underline{x}_j - \underline{\overline{\boldsymbol{x}}})(\underline{x}_j - \underline{\overline{\boldsymbol{x}}})' + \sum_{j=1}^{n}(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})'\right)\right]$$

$$= \operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{B} + n(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})'\right)\right]$$

$$\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left(n\boldsymbol{S}_n + n(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})'\right)\right], \quad \text{con}, \quad \boldsymbol{B} = \sum_{j=1}^{n}(\underline{x}_j - \underline{\overline{\boldsymbol{x}}})(\underline{x}_j - \underline{\overline{\boldsymbol{x}}})' = n\boldsymbol{S}_n$$

La f.d.p conjunta de $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_n$ y función de verosimil

Reemplazando la expresión anterior en la f.d.p.c, se obtiene que:

$$f\left(\underline{\boldsymbol{x}}_1, \underline{\boldsymbol{x}}_2, \cdots, \underline{\boldsymbol{x}}_n \,\middle|\, \underline{\boldsymbol{\mu}}\ ,\ \boldsymbol{\Sigma}\right) =$$

$$\frac{1}{(2\pi)^{np/2}} \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} e^{-\frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}\left[\sum_{j=1}^{n}(\underline{\boldsymbol{x}}_j - \overline{\boldsymbol{x}})(\underline{\boldsymbol{x}}_j - \overline{\boldsymbol{x}})' + n(\overline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}})(\overline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}})'\right]\right\}}$$

y la **función de verosimilitud**, que se obtiene al reemplazar los valores observados de la muestra: $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ en la f.d.p.c anterior es:

$$L\left(\underline{\boldsymbol{\mu}}\ ,\ \boldsymbol{\Sigma} \,\middle|\, \underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n\right) =$$

$$\frac{1}{(2\pi)^{np/2}} \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} e^{-\frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}\left[\sum_{j=1}^{n}(\underline{x}_j - \overline{\boldsymbol{x}})(\underline{x}_j - \overline{\boldsymbol{x}})' + n(\overline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}})(\overline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}})'\right]\right\}}$$

y con $\boldsymbol{B} = \sum_{j=1}^{n} (\underline{x}_j - \underline{\overline{\boldsymbol{x}}})(\underline{x}_j - \underline{\overline{\boldsymbol{x}}})' = n\boldsymbol{S}_n,$ se tiene que:

$$L\left(\underline{\boldsymbol{\mu}}\ ,\ \boldsymbol{\Sigma}\ \middle|\ \underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n\right) = \frac{1}{(2\pi)^{np/2}} \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} e^{-\frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{B} + n(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})'\right]\right\}}$$

$$= \frac{1}{(2\pi)^{np/2}} \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} e^{-\frac{1}{2}\left\{\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{B}\right] + n\mathrm{tr}\left[\underbrace{\boldsymbol{\Sigma}^{-1}(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})}(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})'\right]\right\}}$$

$$= \frac{1}{(2\pi)^{np/2}} \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} e^{-\frac{1}{2}\left\{\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{B}\right] + n\mathrm{tr}\left[(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})'\boldsymbol{\Sigma}^{-1}(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})\right]\right\}}$$

$$L\left(\underline{\boldsymbol{\mu}}\ ,\ \boldsymbol{\Sigma}\right) = \frac{1}{(2\pi)^{np/2}} \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} e^{-\frac{1}{2}\left\{\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{B}\right] + n(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})'\boldsymbol{\Sigma}^{-1}(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})\right\}}$$

Teorema-2: Sea $(\underline{\boldsymbol{x}}_1, \underline{\boldsymbol{x}}_2, \ldots, \underline{\boldsymbol{x}}_n)$ una muestra aleatoria de una población normal p-variada con vector de medias $\underline{\boldsymbol{\mu}}$ y matriz de var-cov $\boldsymbol{\Sigma}$, entonces
los estimadores de máxima-verosimilitud de $\underline{\boldsymbol{\mu}}$ y $\boldsymbol{\Sigma}$ están dados por:

$$\hat{\underline{\boldsymbol{\mu}}} = \overline{\underline{\boldsymbol{x}}} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{bmatrix}_{p\times 1} = \frac{1}{n}\boldsymbol{X}^T\mathbf{1} = \frac{1}{n}\begin{bmatrix} \sum_{i=1}^n x_{i1} \\ \sum_{i=1}^n x_{i2} \\ \vdots \\ \sum_{i=1}^n x_{ip} \end{bmatrix} = \frac{1}{n}\sum_{j=1}^n \underline{\boldsymbol{x}}_j \quad \text{y}$$

$$\hat{\boldsymbol{\Sigma}} = \boldsymbol{S}_n = \frac{\boldsymbol{B}}{n} = \frac{1}{n}\sum_{j=1}^n (\underline{\boldsymbol{x}}_j - \overline{\underline{\boldsymbol{x}}})(\underline{\boldsymbol{x}}_j - \overline{\underline{\boldsymbol{x}}})' = \left(\frac{n-1}{n}\right)\boldsymbol{S}$$

Los valores observados:

$$\overline{\underline{\boldsymbol{x}}}_{p\times 1} = \frac{1}{n}\sum_{j=1}^n \underline{x}_j \quad \text{y} \quad \boldsymbol{S}_n = \frac{1}{n}\sum_{j=1}^n (\underline{x}_j - \overline{\underline{\boldsymbol{x}}})(\underline{x}_j - \overline{\underline{\boldsymbol{x}}})' = \frac{\boldsymbol{B}}{n},$$

son las estimaciones de máxima-verosimilitud de $\underline{\boldsymbol{\mu}}$ y $\boldsymbol{\Sigma}$ respectivamente.

Demostración-1 del teorema anterior:

La función de verosimilitud es:

$$L\left(\underline{\boldsymbol{\mu}}\,,\ \boldsymbol{\Sigma}\ \Big|\ \underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n\right) = \frac{1}{(2\pi)^{np/2}} \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} \mathrm{e}^{-\frac{1}{2}\left\{\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{B}\right] + n(\underline{\overline{\boldsymbol{x}}}-\underline{\boldsymbol{\mu}})'\boldsymbol{\Sigma}^{-1}(\underline{\overline{\boldsymbol{x}}}-\underline{\boldsymbol{\mu}})\right\}}$$

$$\alpha\ \ \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} \mathrm{e}^{-\frac{1}{2}\left\{\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{B}\right] + n(\underline{\overline{\boldsymbol{x}}}-\underline{\boldsymbol{\mu}})'\boldsymbol{\Sigma}^{-1}(\underline{\overline{\boldsymbol{x}}}-\underline{\boldsymbol{\mu}})\right\}}$$

$$L\left(\underline{\boldsymbol{\mu}}\,,\ \boldsymbol{\Sigma}\ \right)\alpha\ \ \frac{1}{|\boldsymbol{\Sigma}|^{n/2}}\ \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{B}\right]} \times \mathrm{e}^{-\frac{1}{2}n(\underline{\overline{\boldsymbol{x}}}-\underline{\boldsymbol{\mu}})'\boldsymbol{\Sigma}^{-1}(\underline{\overline{\boldsymbol{x}}}-\underline{\boldsymbol{\mu}})}$$

esta función se maximiza si su exponente es mínimo, es decir cuando:

$$n(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})'\boldsymbol{\Sigma}^{-1}(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})$$

se minimice, y ésto es mínimo (ie, $L(\ \underline{\boldsymbol{\mu}}\ ,\ \boldsymbol{\Sigma}\ )$ es máxima) con respecto a $\underline{\boldsymbol{\mu}}$, cuando $\underline{\hat{\boldsymbol{\mu}}} = \underline{\overline{\boldsymbol{x}}}$, ie, el MLE de $\underline{\boldsymbol{\mu}}$ es $\underline{\overline{\boldsymbol{x}}}$.

Ahora, resta maximizar a:

$$L(\underline{\hat{\boldsymbol{\mu}}}, \boldsymbol{\Sigma}) \;\; \alpha \;\; \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} e^{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{B}\right]}$$

con respecto a $\boldsymbol{\Sigma}$, donde: $\boldsymbol{B} = \sum_{j=1}^{n}(\underline{x}_j - \underline{\boldsymbol{x}})(\underline{x}_j - \underline{\boldsymbol{x}})' = n\boldsymbol{S}_n$.

Para obtener el MLE de $\boldsymbol{\Sigma}$, es decir, para maximizar la función $L(\underline{\hat{\boldsymbol{\mu}}}, \boldsymbol{\Sigma})$ con respecto a $\boldsymbol{\Sigma}$, se utilizará el siguiente teorema.

Teorema: Dada una matríz simétrica definida positiva $\boldsymbol{B}_{p\times p}$ y un escalar $b > 0$, entonces:

$$\frac{1}{|\boldsymbol{\Sigma}|^{b}} e^{-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{B}\right)} \leq \frac{1}{|\boldsymbol{B}|^{b}}(2b)^{(pb)} e^{-bp} \, , \quad \forall \; \text{matriz } \; \boldsymbol{\Sigma} \; \text{ Def } +,$$

y la igualdad se cumple sólo cuando: $\boldsymbol{\Sigma} = \left(\frac{1}{2b}\right)\boldsymbol{B}$.

ie. el máximo de: $\displaystyle\frac{1}{|\boldsymbol{\Sigma}|^{b}} e^{-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{B}\right)}$

se alcanza en: $\boldsymbol{\Sigma} = \left(\frac{1}{2b}\right)\boldsymbol{B}$.

Usando el teorema anterior, con:

$$b = \frac{n}{2} \quad \text{y} \quad \boldsymbol{B} = \sum_{j=1}^{n} (\underline{x}_j - \underline{\overline{\boldsymbol{x}}})(\underline{x}_j - \underline{\overline{\boldsymbol{x}}})',$$

se obtiene que el máximo de:

$$L(\hat{\underline{\boldsymbol{\mu}}}\ ,\ \boldsymbol{\Sigma}) \ \alpha \ \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} e^{-\frac{1}{2}\mathrm{tr}[\boldsymbol{\Sigma}^{-1}\boldsymbol{B}]}$$

ocurre cuando:

$$\hat{\boldsymbol{\Sigma}} = \left(\frac{1}{2b}\right)\boldsymbol{B} = \frac{1}{n}\sum_{j=1}^{n} (\underline{x}_j - \underline{\overline{\boldsymbol{x}}})(\underline{x}_j - \underline{\overline{\boldsymbol{x}}})' = \boldsymbol{S}_n = \left(\frac{n-1}{n}\right)\boldsymbol{S}$$

En resumen, los estimadores MLE de $\underline{\boldsymbol{\mu}}$ y $\boldsymbol{\Sigma}$ son:

El vector de medias muestrales: $\hat{\underline{\boldsymbol{\mu}}} = \underline{\overline{\boldsymbol{x}}}$

y la matriz de var-cov muestrales:

$$\hat{\boldsymbol{\Sigma}} = \boldsymbol{S}_n = \frac{\boldsymbol{B}}{n} = \frac{1}{n}\sum_{j=1}^{n} (\underline{x}_j - \underline{\overline{\boldsymbol{x}}})(\underline{x}_j - \underline{\overline{\boldsymbol{x}}})' = \left(\frac{n-1}{n}\right)\boldsymbol{S}$$

Observaciones finales:

Notar que: El vector de medias muestrales $\underline{\hat{\mu}} = \underline{\overline{x}}$: es un vector aleatorio y la matriz de var-cov muestrales $\hat{\mathbf{\Sigma}} = \mathbf{S}_n = \frac{1}{n}\sum_{j=1}^{n}(\underline{x}_j - \underline{\overline{x}})(\underline{x}_j - \underline{\overline{x}})'$: también es una matriz aleatoria, ya que sus valores cambian de muestra a muestra.

Las estimaciones de máx-verosimilitud son sus respectivos valores particulares obtenidos para el conjunto particular de datos u observaciones dadas.

Ahora Como,

$$\frac{1}{|\mathbf{\Sigma}|^b} e^{-\frac{1}{2}\operatorname{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{B}\right)} \leq \frac{1}{|\mathbf{B}|^b}(2b)^{(pb)} e^{-bp}$$

El valor máximo de la función de verosimilitud $L\left(\underline{\mu}\,,\ \mathbf{\Sigma}\right)$ es:

$$L\left(\underline{\hat{\boldsymbol{\mu}}}\ ,\ \hat{\boldsymbol{\Sigma}}\right) \ = \ \frac{1}{|\boldsymbol{B}|^{b}}(2b)^{(pb)}e^{-bp} \quad ; \quad \text{con} \quad \boldsymbol{B} = \hat{\boldsymbol{\Sigma}}(2b), \quad \text{y} \quad b = \frac{n}{2}$$

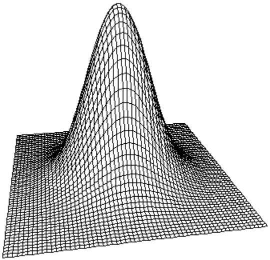
$$= \frac{1}{(2\pi)^{np/2}} \frac{1}{(2b)^{np/2}|\hat{\boldsymbol{\Sigma}}|^{n/2}}(2b)^{np/2}e^{-\frac{np}{2}},$$

$$= \underbrace{\frac{1}{(2\pi)^{np/2}}e^{-\frac{np}{2}}\left(\frac{n-1}{n}\right)^{-\frac{np}{2}}}|\boldsymbol{S}|^{-\frac{n}{2}}$$

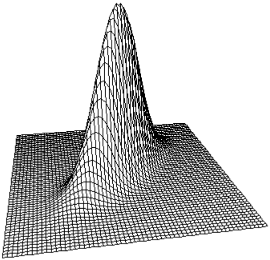
$$= \text{k}|\boldsymbol{S}|^{-\frac{n}{2}} = \text{k}(\text{ Varianza Generalizada})^{-\frac{n}{2}}$$

$$L(\ \underline{\hat{\boldsymbol{\mu}}}\ ,\ \hat{\boldsymbol{\Sigma}}\ ) = \frac{\text{k}}{(\text{ Varianza Generalizada})^{\frac{n}{2}}}$$

De acuerdo a lo anterior, la **Varianza-Generalizada** determina lo " impinada" o "no-impinada" que es la función de verosimilitud para "un conjunto o muestra de datos particular", por lo que es una medida-natural de **variabilidad** cuando la población es normal multivariada, entre más grande mas achatada ie. más variabilidad y entre mas pequeña mas impinada ie. menos variabilidad.



Más Grande
Más Achatada, Más Variabilidad



Más Pequeña
Más Impinada, Menos Variabilidad

Propiedades de Invarianza de los MLE

Sea $\hat{\underline{\boldsymbol{\theta}}}$ el MLE de una parámetro multivariado $\underline{\boldsymbol{\theta}}$ y se considera la estimación del $h(\underline{\boldsymbol{\theta}})$, lo cual es una función de $\underline{\boldsymbol{\theta}}$, entonces el MLE de $h(\underline{\boldsymbol{\theta}})$ está dado por: $h(\hat{\underline{\boldsymbol{\theta}}})$.

Ejemplos:

a) El MLE de $\underline{\boldsymbol{\mu}}'\boldsymbol{\Sigma}^{-1}\underline{\boldsymbol{\mu}}$ es: $\hat{\underline{\boldsymbol{\mu}}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\underline{\boldsymbol{\mu}}} = \overline{\boldsymbol{x}}\boldsymbol{S}_n^{-1}\overline{\boldsymbol{x}}$.

b) El MLE de $\sqrt{\sigma_{ii}}$ es $\sqrt{\hat{\sigma}_{ii}}$, en donde:

$$\hat{\sigma}_{ii} = \frac{1}{n}\sum_{j=1}^{n}(X_{ji} - \overline{X}_i)^2 = \left(\frac{n-1}{n}\right)s_{ii}, \text{ es el MLE de } \sigma_{ii} = Var(X_i).$$

Demostración-2 del Teorema Anterior

Teorema-4: Sea $(\underline{\boldsymbol{x}}_1, \underline{\boldsymbol{x}}_2, \ldots, \underline{\boldsymbol{x}}_n)$ una muestra aleatoria de una población normal p-variada con vector de medias $\underline{\boldsymbol{\mu}}$ y matriz de var-cov $\boldsymbol{\Sigma}$, entonces los estimadores de máxima verosimilitud de $\underline{\boldsymbol{\mu}}$ y $\boldsymbol{\Sigma}$ están dados por:

$$\underline{\hat{\boldsymbol{\mu}}} = \underline{\overline{\boldsymbol{x}}} \;\; \text{y} \quad \hat{\boldsymbol{\Sigma}} = \boldsymbol{S}_n = \frac{\boldsymbol{B}}{n} = \left(\frac{n-1}{n}\right)\boldsymbol{S}$$

Demostración: La función de verosimilitud, esta dada por:

$$L\Big(\underline{\boldsymbol{\mu}} \,,\, \boldsymbol{\Sigma} \;\Big|\; \underline{x}_1, \ldots, \underline{x}_n \Big) =$$

$$\frac{1}{(2\pi)^{np/2}} \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} e^{-\frac{1}{2}\text{tr}\left\{\boldsymbol{\Sigma}^{-1}\left[\sum_{j=1}^{n}(\underline{x}_j-\underline{\overline{\boldsymbol{x}}})(\underline{x}_j-\underline{\overline{\boldsymbol{x}}})'+n(\underline{\overline{\boldsymbol{x}}}-\underline{\boldsymbol{\mu}})(\underline{\overline{\boldsymbol{x}}}-\underline{\boldsymbol{\mu}})'\right]\right\}}$$

y con $\boldsymbol{B} = \sum_{j=1}^{n}(\underline{x}_j - \underline{\overline{\boldsymbol{x}}})(\underline{x}_j - \underline{\overline{\boldsymbol{x}}})' = n\boldsymbol{S}_n$, se tiene que:

$$L\Big(\underline{\boldsymbol{\mu}} \,,\, \boldsymbol{\Sigma} \Big) = \frac{1}{(2\pi)^{np/2}} \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} e^{-\frac{1}{2}\text{tr}\left\{\boldsymbol{\Sigma}^{-1}\left[n\boldsymbol{S}_n+n(\underline{\overline{\boldsymbol{x}}}-\underline{\boldsymbol{\mu}})(\underline{\overline{\boldsymbol{x}}}-\underline{\boldsymbol{\mu}})'\right]\right\}}$$

Tomando logaritmo se tiene que la función log de verosimilitud es:

$$l(\underline{\boldsymbol{\mu}}\ ,\ \boldsymbol{\Sigma}) = -\frac{np}{2}\mathrm{Log}(2\pi) + \frac{n}{2}\underbrace{\mathrm{Log}\left(|\boldsymbol{\Sigma}^{-1}|\right)} - \frac{n}{2}\underbrace{\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{S}_n\right]}$$

$$-\frac{n}{2}\underbrace{\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})'\right]} \quad ; \quad |\boldsymbol{\Sigma}^{-1}| = \frac{1}{|\boldsymbol{\Sigma}|}$$

Ahora, $\quad l\left(\ \underline{\boldsymbol{\mu}}\ ,\ \boldsymbol{\Sigma}\ \right)$

$$= -\frac{np}{2}\mathrm{Log}(2\pi) + \frac{n}{2}\mathrm{Log}\left(|\boldsymbol{\Sigma}^{-1}|\right) - \frac{n}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{S}_n\right] - \frac{n}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})'\right]$$

$$= -\frac{np}{2}\mathrm{Log}(2\pi) + \frac{n}{2}\mathrm{Log}\left(|\boldsymbol{\Sigma}^{-1}|\right) - \frac{n}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{S}_n\right] - \frac{n}{2}\left[(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})^t\boldsymbol{\Sigma}^{-1}(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})\right]$$

Para derivar parcialmente a $l(\underline{\boldsymbol{\mu}}\ ,\ \boldsymbol{\Sigma})$, con respecto a $\underline{\boldsymbol{\mu}}$ y $\boldsymbol{\Sigma}^{-1}$, se usarán los siguientes resultados:

1. $\frac{\partial(\underline{\boldsymbol{x}}^t \boldsymbol{A}\underline{\boldsymbol{x}})}{\partial \underline{\boldsymbol{x}}} = 2\boldsymbol{A}\underline{\boldsymbol{x}}$. Para $\boldsymbol{A}$-Simétrica. Con: $\underline{\boldsymbol{x}} = (\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})$ y $\boldsymbol{A} = \boldsymbol{\Sigma}^{-1}$
2. La derivada del logaritmo del determinante de una matriz $\boldsymbol{X}_{p\times p}$ es:
$$\frac{\partial \mathrm{Log}(|\boldsymbol{X}|)}{\partial \boldsymbol{X}} = \frac{1}{|\boldsymbol{X}|}\frac{\partial |\boldsymbol{X}|}{\partial \boldsymbol{X}} = 2\boldsymbol{X}^{-1} - \mathrm{Diag}(\boldsymbol{X}^{-1})$$
Con: $\boldsymbol{X} = \boldsymbol{\Sigma}^{-1}$.
3. $\frac{\partial[tr(\boldsymbol{X}\boldsymbol{A})]}{\partial \boldsymbol{X}} = \boldsymbol{A} + \boldsymbol{A}' - \mathrm{Diag}(\boldsymbol{A})$, si $\boldsymbol{X}$-es simétrica. Con:
$\boldsymbol{X} = \boldsymbol{\Sigma}^{-1}$ y $\boldsymbol{A} = \boldsymbol{S}_n$

Aplicando los tres resultados anteriores se tiene que:

$$\frac{\partial l}{\partial \underline{\boldsymbol{\mu}}} = n\boldsymbol{\Sigma}^{-1}(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})$$

$$\frac{\partial \mathrm{Log}(|\boldsymbol{\Sigma}^{-1}|)}{\partial \boldsymbol{\Sigma}^{-1}} = 2\boldsymbol{\Sigma} - \mathrm{Diag}(\boldsymbol{\Sigma})$$

$$\frac{\partial \mathrm{tr}[\boldsymbol{\Sigma}^{-1}\boldsymbol{S}_n]}{\partial \boldsymbol{\Sigma}^{-1}} = 2\boldsymbol{S}_n - \mathrm{Diag}(\boldsymbol{S}_n)$$

$$\frac{\partial \mathrm{tr}[\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}} - \underline{\boldsymbol{\mu}})(\overline{\mathbf{x}} - \underline{\boldsymbol{\mu}})']}{\partial \boldsymbol{\Sigma}^{-1}} = 2(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})' - \mathrm{Diag}[(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})']$$

luego, igualando a cero las derivadas parciales: $\frac{\partial l}{\partial \underline{\boldsymbol{\mu}}}$ y $\frac{\partial l}{\partial \boldsymbol{\Sigma}^{-1}}$, se tiene que:

$$n\boldsymbol{\Sigma}^{-1}(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}}) = \underline{\mathbf{0}}$$

$$\frac{n}{2}\left[2\boldsymbol{\Sigma} - \mathrm{Diag}(\boldsymbol{\Sigma})\right] - \frac{n}{2}\left[2\boldsymbol{S}_n - \mathrm{Diag}(\boldsymbol{S}_n)\right] - \frac{n}{2}\left\{2(\underline{\overline{\boldsymbol{x}}}-\underline{\boldsymbol{\mu}})(\underline{\overline{\boldsymbol{x}}}-\underline{\boldsymbol{\mu}})' - \mathrm{Diag}\left[(\underline{\overline{\boldsymbol{x}}}-\underline{\boldsymbol{\mu}})(\underline{\overline{\boldsymbol{x}}}-\underline{\boldsymbol{\mu}})'\right]\right\} = 0$$

De la primera igualdad se obtiene que: $\underline{\hat{\boldsymbol{\mu}}} = \underline{\overline{\boldsymbol{x}}}$

y de la segunda igualdad se tiene que:

$$\frac{n}{2}\left[2\boldsymbol{\Sigma} - 2\boldsymbol{S}_n - 2(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})^{'}\right] - \frac{n}{2}\left[\mathrm{Diag}(\boldsymbol{\Sigma}) - \mathrm{Diag}(\boldsymbol{S}_n) - \mathrm{Diag}[(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})^{'}]\right] = 0$$

es decir,

$$\left[\boldsymbol{\Sigma} - \boldsymbol{S}_n - (\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})^{'}\right] - \frac{1}{2}\mathrm{Diag}\left[\boldsymbol{\Sigma} - \boldsymbol{S}_n - (\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})^{'}\right] = 0$$

de donde,

$$\left[\boldsymbol{\Sigma} - \boldsymbol{S}_n - (\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})^{'}\right] = \frac{1}{2}\mathrm{Diag}\left[\boldsymbol{\Sigma} - \boldsymbol{S}_n - (\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})^{'}\right]$$

y por lo tanto:

$$\boldsymbol{\Sigma} - \boldsymbol{S}_n - (\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})(\underline{\overline{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})^{'} = 0$$

es decir: $\hat{\boldsymbol{\Sigma}} = \boldsymbol{S}_n + (\underline{\overline{\boldsymbol{x}}} - \underline{\hat{\boldsymbol{\mu}}})(\underline{\overline{\boldsymbol{x}}} - \underline{\hat{\boldsymbol{\mu}}})^{'}$ y como $\underline{\hat{\boldsymbol{\mu}}} = \underline{\overline{\boldsymbol{x}}}$, entonces

$$\hat{\boldsymbol{\Sigma}} = \boldsymbol{S}_n \;, \quad l.q.q.d.$$

Distribución Muestral de $\underline{\overline{x}}$ y $\boldsymbol{S}_n$

Caso univariado

Recordemos del caso univariado, que si si $p = 1$, entonces:
$\overline{X} \sim N\Big(\mu\ ,\ \frac{\sigma^2}{n}\Big)$, es decir que,

$$\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n} = \frac{\text{Varianza Poblacional}}{\text{Tamaño Muestral}}.$$

Para el caso de S^2, recordemos el siguiente resultado del caso univariado:

$$\frac{nS_n^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}, \qquad ie. \qquad X^2 = (n-1)S^2 \sim \sigma^2\chi^2_{(n-1)}$$

con: $\quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$

Caso Multivariado

En el caso multivariado, ie, $p \geq 2$, el resultado es similar para $\underline{\bar{\boldsymbol{x}}}$, es decir que:

$$\underline{\bar{\boldsymbol{x}}} \sim N_p\left(\underline{\boldsymbol{\mu}}\ ,\ \frac{\boldsymbol{\Sigma}}{n}\right), \qquad ie. \qquad \mathrm{Var}\left(\underline{\bar{\boldsymbol{x}}}\right) = \frac{\boldsymbol{\Sigma}}{n} = \frac{\text{Varianza Poblacional}}{\text{Tamaño Muestral}}.$$

Para $\boldsymbol{S}$, se tiene la **Distribución Matriz Variada Wishart** o simplemente **Distribución-Wishart** , la cual se define como una suma de productos de vectores-aleatorios normales multivariados.

Definición-1: Sean $\underline{\boldsymbol{z}}_j$ vectores aleatorios p-variados independientes e identicamente distribuídos normales multivariados con vector de medias $\underline{\boldsymbol{\mu}} = \mathbf{0}$ y matriz de var-cov, $\boldsymbol{\Sigma}$, es decir, $\underline{\boldsymbol{z}}_j \sim N_p(\mathbf{0}\ ,\ \boldsymbol{\Sigma})$, entonces la matriz aleatoria $\boldsymbol{W}_{p\times p} = \sum_{j=1}^{n} \underline{\boldsymbol{z}}_j \underline{\boldsymbol{z}}_j^{'}$ se dice que tiene una distribución matriz-variada Wishart con parámetros $\boldsymbol{\Sigma}$ y n, y se denota por:

$$\boldsymbol{W}_{p\times p} = \sum_{j=1}^{n} \underline{\boldsymbol{z}}_j \underline{\boldsymbol{z}}_j^{'} \sim W_p\left(\boldsymbol{\Sigma}\ ,\ n\right)$$

Teorema-5: Distribución de $\underline{\overline{\boldsymbol{x}}}$ y $\boldsymbol{S}_n$

Sea $\underline{\boldsymbol{x}}_1, \underline{\boldsymbol{x}}_2, \ldots, \underline{\boldsymbol{x}}_n$ una muestra aleatoria de tamaño n de una distribución normal p-variada con media $\underline{\boldsymbol{\mu}}$ y matriz de var-cov $\boldsymbol{\Sigma}$, ie, $\underline{\boldsymbol{x}}_j \sim N_p(\underline{\boldsymbol{\mu}}\ ,\ \boldsymbol{\Sigma})$, entonces:

$$\underline{\overline{\boldsymbol{x}}} \sim N_p\left(\underline{\boldsymbol{\mu}}\ ,\ \frac{\boldsymbol{\Sigma}}{n}\right) \qquad \text{y} \qquad \boldsymbol{W} = n\boldsymbol{S}_n = (n-1)\boldsymbol{S} \sim W_p\Big(\boldsymbol{\Sigma}\ ,\ n-1\Big)$$

Además, $\underline{\overline{\boldsymbol{x}}}$ y $\boldsymbol{S}$ son independientes.

Función de densidad de probabilidad de la Wishart:

La f.d.p de una matriz aleatoria distribuída Wishart, ie, si $\boldsymbol{A} \sim W_p\Big(\boldsymbol{\Sigma}\ ,\ n\Big)$, entonces su f.d.p se puede expresar de forma explícita, en el caso $n > p$, como sigue:

$$f_W(\ A \mid \boldsymbol{\Sigma}\ ,\ n) = \frac{|A|^{\frac{n-p-2}{2}}\, e^{-\mathrm{tr}[A\boldsymbol{\Sigma}^{-1}]^{1/2}}}{2^{\frac{(n-1)p}{2}}\pi^{\frac{p(p-1)}{4}}|\boldsymbol{\Sigma}|^{\frac{n-1}{2}}\prod_{i=1}^{n}\boldsymbol{\Gamma}\left(\frac{n-i}{2}\right)}$$

Algunas Propiedades de la Distribución Wishart

1. Si $\boldsymbol{W} \sim W_p\left(\boldsymbol{\Sigma}\ ,\ n\right)$ y $\boldsymbol{A}_{p\times p}$ es una matriz de constantes, entonces,

$$\boldsymbol{A}^{'}\boldsymbol{W}\boldsymbol{A} \sim W_p\left(\boldsymbol{A}^{'}\boldsymbol{\Sigma}\boldsymbol{A}\ ,\ n\right)$$

2. Si $\boldsymbol{W} \sim W_p\left(\boldsymbol{\Sigma}\ ,\ n\right)$ y $\underline{a} \in \mathbb{R}^p$ con $\underline{a}^{'}\underline{a} \neq 0$, entonces,

$$\frac{\underline{a}^{'}\boldsymbol{W}\underline{a}}{\underline{a}^{'}\underline{a}} \sim \chi^2_{(n)}$$

3. Si $\boldsymbol{W}_1 \sim W_p\left(\boldsymbol{\Sigma}\ ,\ n_1\right)$ y $\boldsymbol{W}_2 \sim W_p\left(\boldsymbol{\Sigma}\ ,\ n_2\right)$, con $\boldsymbol{W}_1$ y $\boldsymbol{W}_2$ independientes, entonces,

$$\boldsymbol{W}_1 + \boldsymbol{W}_2 \sim W_p\left(\boldsymbol{\Sigma}\ ,\ n_1 + n_2\right)$$

4. Si $\boldsymbol{C}_{n\times n}$: es una matriz simétrica e idempotente y $\boldsymbol{X}_{n\times p}$ es una matriz de datos de una $N_p(\boldsymbol{0}\ ,\ \boldsymbol{\Sigma})$, entonces,

$$\boldsymbol{X}^{'}\boldsymbol{C}\boldsymbol{X} \sim W_p\left(\boldsymbol{\Sigma}\ ,\ r\right), \quad \text{con} \quad r = \text{tr}(C)$$

Definición-2: T^2-de Hotelling.

Sea $\underline{\boldsymbol{x}}_p$ un vector aleatorio tal que, $\underline{\boldsymbol{x}} \sim N_p(\underline{\boldsymbol{\mu}}\ ,\ \boldsymbol{\Sigma})$ y $\boldsymbol{W} \sim W_p(\boldsymbol{\Sigma}\ ,\ n)$, donde $\underline{\boldsymbol{x}}$ y $\boldsymbol{W}$ son independientes, con $\boldsymbol{\Sigma}$ y $\boldsymbol{W}$ son No-singulares.

Se le llama **Estadístico T^2–de Hotelling** al estadístico definido como:

$$T^2 = n(\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}})' \boldsymbol{W}^{-1} (\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}})$$

A la distribución del estadístico T^2 se le conoce como la **Distribución T^2-de Hotelling** y se denota por:

$$T^2 \sim T^2\Big(p\ ,\ n\Big).$$

Caso aprticular, para: $\boldsymbol{\Sigma} = (n-1)\boldsymbol{S} \sim W_p(p, n-1)$, entonces se tien que:

$$T^2 = (n-1)(\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}})' \boldsymbol{S}^{-1} (\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}}) \sim T^2\Big(p\ ,\ n-1\Big).$$

Teorema-6: Suponga que $\underline{\boldsymbol{x}} \sim N_p(\underline{\boldsymbol{\mu}}\ ,\ \boldsymbol{\Sigma})$ y $\boldsymbol{W} \sim W_p(\boldsymbol{\Sigma}\ ,\ n)$, donde $\underline{\boldsymbol{x}}$ y $\boldsymbol{W}$ son independientes, con $\boldsymbol{\Sigma}$ y $\boldsymbol{W}$ son No-singulares, entonces:

$$F = \frac{(n-p+1)}{np} T^2 \sim F_{p\ ,\ n-p+1}$$

Comportamiento de $\underline{\overline{x}}_p$ y S para tamaños muestrales grandes

Ley Débil de los Grandes Números:

Caso Univariado:

Si $X_1, X_2, \ldots, X_n$ es una m.a de tamaño n de una población con media

$E[X_i] = \mu$, entonces

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} \xrightarrow{p} \mu,$$

Caso Multivariado:

Sea $\underline{\boldsymbol{x}}_1, \underline{\boldsymbol{x}}_2, \ldots, \underline{\boldsymbol{x}}_n$ es una m.a de tamaño n de una población con distribución p-variada con vector de medias $E[\underline{\boldsymbol{x}}_i] = \underline{\boldsymbol{\mu}}$ y matriz de var-cov $\boldsymbol{\Sigma}$.

Si $\bar{\underline{\boldsymbol{x}}} = \frac{1}{n}\sum_{i=1}^{n} \underline{\boldsymbol{x}}_i$, es el vector de medias muestrales, entonces,

$$\bar{\underline{\boldsymbol{x}}} \xrightarrow{p} \underline{\boldsymbol{\mu}},$$

Lo anterior implica que:

$$\overline{X}_i \xrightarrow{p} \mu_i \ , \ \text{para} \ \ i = 1, 2, \ldots, p.$$

De manera similar, se tiene que:

$$\boldsymbol{S}_n \xrightarrow{p} \boldsymbol{\Sigma},$$

Lo anterior implica que:

$$S_{ij} \xrightarrow{p} \sigma_{ij} \ , \ \text{para} \ \ i, j = 1, 2, \ldots, p.$$

Teorema Central del Límite (TCL).

Caso Univariado:

Si $X_1, X_2, \ldots, X_n$ es una m.a de tamaño n de una población con media

$E[X_i] = \mu$ y varianza σ^2, entonces,

$$\sqrt{n}(\overline{X} - \mu) \xrightarrow{d} N(0\ ,\ \sigma^2),$$

y por lo tanto,

$$\frac{\sqrt{n}(\overline{X} - \mu)}{\sigma} \xrightarrow{d} N(0\ ,\ 1)$$

Caso Multivariado:

Sea $\underline{\boldsymbol{x}}_1, \underline{\boldsymbol{x}}_2, \ldots, \underline{\boldsymbol{x}}_n$ es una m.a de tamaño n de una población con distribución p-variada con vector de medias $E[\underline{\boldsymbol{x}}_i] = \underline{\boldsymbol{\mu}}$ y matriz de var-cov $\boldsymbol{\Sigma}$, entonces,

$$\underline{\boldsymbol{z}} = \sqrt{n}(\underline{\bar{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}}) \xrightarrow{d} N_p\Big(\boldsymbol{0}\ ,\ \boldsymbol{\Sigma}\Big)$$

Además, si $\boldsymbol{\Sigma} > 0$, entonces,

$$\underline{\tilde{\boldsymbol{z}}} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\sqrt{n}(\underline{\bar{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}}) \xrightarrow{d} N_p\Big(\boldsymbol{0}\ ,\ \boldsymbol{I}_p\Big), \quad \text{y}$$

$$n(\underline{\bar{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})' \boldsymbol{S}^{-1}(\underline{\bar{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}}) \xrightarrow{d} \chi^2_p,$$

es decir, converge aproximadamente a una chi-cuadrado, cuando $n - p >>$.

Algunas funciones del R para Análisis Multivariado

Explorar las siguientes Librerías y Funciones del R para Análisis de Datos Multivariados.

- Paquete mvn, mva
- Se Encuentran 27 librerías con funciones para análisis de datos multivariados:

  ade4 amap cclust culster CoCoAn DAAG fpc gclus
  hier.part knncat knnTree multidim multiv vnormtest
  mvpart mvtnorm norm pan pcurve prabclus princurve
  rpart sca tree vegan xgobi scatterplot3d

  ‘ade4 amap cclust culster CoCoAn DAAG fpc gclus hier.part knncat knnTree multidim multiv vnormtest mvpart mvtnorm norm pan pcurve prabclus princurve rpart sca tree vegan xgobi scatterplot3d‘

- ADE4 231 items (60 ficheros de datos): 171 funciones
- Tenemos points(), lines(), segments(), rect(), arrows(), text(), abline(), . . .
- Algunas funciones importantes:

```
kmeans() dist() hclust() cov() cor() scale() table()
cutree() acp() acb() cmdscale() mca() biplot()
dibujaacp() dibujaacb()
```

‘kmeans() dist() hclust() cov() cor() scale() table() cutree() acp() acb() cmdscale() mca() biplot() dibujaacp() dibujaacb()‘