Chapter 2

Distribución Normal Multivariada

Raúl Alberto Pérez

diciembre 14, 2022

Contents

1 Repaso de álgebra lineal **1**

1.1 Algunos conceptos básicos 1

1.2 Diferenciación con Vectores y Matrices 7

1.2.1 Vector-Gradiente para diferentes definiciones de $f(\underline{x})$: . . 8

1.3 Organización y Presentación de Datos 10

1.3.1 Resumenes Descriptivos 10

1.3.2 Representación Gráfica de Observaciones multivariadas . 11

1.3.3 Vectores y Matrices Aleatorias 11

1.4 Vectores y Matrices Poblacionales Particionados 15

1.4.1 Particionamiento del Vector de Medias-Poblacionales . . . 15

1.4.2 Particionamiento de la Matriz de Var-Cov Poblacional . 15

1.5 Propiedades Sobre la Media y Varianza de Combinaciones Lineales 16

1.6 Vectores y Matrices Muestrales Particionadas 18

1.6.1 Particionamiento del Vector de Medias Muestrales 18

1.6.2 Particionamiento de la Matriz de Var-Cov Muestrales . . 19

1.7 Algunas formas matriciales Eficientes 19

1.8 Muestra Aleatoria de Distribuciones p-Variadas 20

1.9 Distancias 23

1.9.1 Introducción 23

2 Distribución Normal Multivariada **27**

2.1 Geometría y propiedades de la NM 27

2.2 Normal Univariada 27

2.3 Normal Multivariada 28

2.3.1 Algunos Aspectos Geométricos de la NM 28

2.4 Propiedades de la distribución Normal Multivariada 34

2.5 Evaluación del Supuesto de Normalidad Multivariada 52

2.5.1 Evaluación a nivel marginal (ie. Normalidad Univariada) 54

Prefacio

Introducción al Análisis Multivariado se escribió con la intención de …………… usando [R Core Team, 2022] y además……….

Agradecimientos. Quiero agradecer a mis compañeros de trabajo Victor López, ……….

```
# R packages needed for this book
#   along with versions used when book knitted in Jan 2021

# Install CRAN packages needed
needed_pkgs <- c(
  "bookdown",     # 0.19
  "boot",         # 1.3-25
  "broom",        # 0.5.6
  "data.table",   # 1.12.8
  "GGally",       # 2.0.0
  "ggmosaic",     # 0.2.0
  "gridExtra",    # 2.3
  "Hmisc",        # 4.4-0
  "HMLdiag",      # 0.3.1
  "ICC",          # 2.3.0
  "jtools",       # 2.0.5
  "kableExtra",   # 1.1.0
  "knitr",        # 1.28
  "lattice",      # 0.20-41
  "lme4",         # 1.1-23
  "MASS",         # 7.3-51.6
  "mice",         # 3.9.0
  "mnormt",       # 2.0.0
  "mosaic",       # 1.7.0
  "multcomp",     # 1.4-13
  "nlme",         # 3.1-148
  "pander",       # 0.6.3
```

```
  "pscl",         # 1.5.5
  "reshape2",     # 1.4.4
  "rsample",      # 0.0.7
  "sessioninfo",  # 1.1.1
  "tidyverse",    # 1.3.0
  "xtable"        # 1.8-4
)

## Automatically installing needed packages not yet installed
#new_pkgs <- needed_pkgs[!(needed_pkgs %in%
#  installed.packages())]
#if (length(new_pkgs)) {
#  install.packages(new_pkgs, repos = "http://cran.rstudio.com")
#}
```

```
if(knitr::is_html_output()){options(knitr.table.format = "html")} else {options(knitr.t
```

Chapter 1

Repaso de álgebra lineal

1.1 Algunos conceptos básicos

Definición 1.1 (Vector)**.** Arreglo de n-números reales

Definición 1.2 (Vector de Unos)**.** Vector cuyas entradas son todos uno

Definición 1.3 (Suma de Vectores)**.** Se realiza la suma componente a componente

Definición 1.4 (Producto interno entre vectores)**.** Multiplicación elemento a elemento, ie.

$$\langle \underline{a}\ ,\ \underline{b} \rangle = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = \sum_i a_i b_i$$

Definición 1.5 (Norma de un Vector)**.** Es la raíz cuadrada del producto interno del vector por si mismo, ie.

$$\|\underline{a}\| = \sqrt{\langle \underline{a}\ ,\ \underline{a} \rangle} = \sqrt{\underline{a}^t \underline{a}} = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2},$$

con $\underline{a}^t = (a_1, a_2, \dots, a_n)$

Propiedades de la Norma:

$\|c\underline{a}\| = |c|\|\underline{a}\|$, si $|c| > 1$ entonces $\underline{a}$-se extiende y si $|c| < 1$ entonces $\underline{a}$-se contrae.

Definición 1.6 (Distancia entre dos vectores)**.** $d(\underline{a}\ ,\ \underline{b}) = \|\underline{a} - \underline{b}\|$

Definición 1.7 (Ángulo entre 2-vectores)**.** El ángulo θ entre dos vectores $\underline{a}$ y $\underline{b}$ está dado por:

$$\cos\theta = \frac{\langle \underline{a}\ ,\ \underline{b} \rangle}{\|\underline{a}\|\|\underline{b}\|} = \frac{\underline{a}^t \underline{b}}{\|\underline{a}\|\|\underline{b}\|}$$

Definición 1.8 (Vectores ortogonales)**.** $\underline{a}$ y $\underline{b}$ son ortogonales si $\langle \underline{a}\ ,\ \underline{b} \rangle = 0$

Definición 1.9 (Proyección Ortogonal)**.** La proyección ortogonal de un vector $\underline{u}$ sobre un vector $\underline{v}$ es el vector $\underline{u}_p = proy_{\underline{v}}\underline{u}$, definido por:

$$\underline{u}_p = proy_{\underline{v}}\underline{u} = \frac{\langle \underline{u} \ , \ \underline{v} \rangle}{\|\underline{v}\|^2}\underline{v} = k.\underline{v},$$

con $k = \frac{\langle \underline{u} \ , \ \underline{v} \rangle}{\|\underline{v}\|^2}$ y $\|\underline{v}\|^2 = \langle \underline{v} \ , \ \underline{v} \rangle$

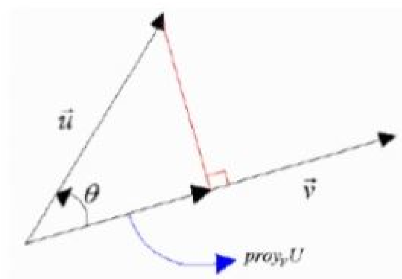

Figure 1.1: Proyección Ortogonal

Definición 1.10 (Matriz)**.** Una matriz es un arreglo rectangular de números en filas y columnas. Usualmente se denotan por $A_{n\times p}$

Definición 1.11 (Matriz Cuadrada)**.** $A_{n\times p}$ es cuadrada si $n = p$, ie. número de filas igual al número de columnas, se denota por: A_n

Definición 1.12 (Matriz Diagonal)**.** Es una matriz cuadrada con ceros fuera de la diagonal, se denota por D_n

Definición 1.13 (Matriz Identidad)**.** Es una matriz diagonal con unos en la diagonal, se denota por I_n

Definición 1.14 (Matriz Cuadrada de unos)**.** Es una matriz cuadrada llena de unos, se denota por: J_n

Definición 1.15 (Matriz triangular inferior y triangular superior)**.** Matrices con parte superior llena de ceros y parte inferior llena de ceros, respectivamente.

Definición 1.16 (Suma de Matrices)**.** La suma de matrices de igual dimensión se realiza componente a componente.

Definición 1.17 (Multiplicación de Matrices)**.** La multiplicación de matrices de dimensiones apropiadas se realiza siguiendo el procedimiento de filas por columnas.

Definición 1.18 (Matriz Traspuesta)**.** La transpuesta de una matriz $A_{n\times p}$ es la matriz que se obtiene invirtiendo las filas por columnas, la cual se denota por: $A^t_{p\times n}$. Se cumple que: $(AB)^t = B^tA^t$.

Definición 1.19 (Matriz Simétrica)**.** Una matriz simétrica es una matriz cuadrada A_n tal que su transpuesta es igual a la matriz original, ie. si $A^t = A$.

Definición 1.20 (Determinante de una Matriz)**.** Sea A una matriz cuadrada de orden n. El determinante de A_n denotado por $|A|$, se obtiene como:

$$|A| = \sum_{j=1}^{n} a_{ij}|A_{ij}|(-1)^{i+j},$$

en donde, A_{ij}-es la matriz cuadrada de orden $(n-1)$ que se obtiene al eliminar la fila i y la columna j de A (ie. menor A_{ij}).

Alguna propiedades del determinante de una matriz:

Sean A y B matrices cuadradas de orden n y $c \in \mathbb{R}$, entonces:

a. $|A| = |A^t|$

b. $|AB| = |A||B|$

c. $|cA| = c^n|A|$

d. La matriz A es invertible si el $|A| \neq 0$, esto equivale a decir que todos los valores propios de A son diferentes de cero.

e. Si la inversa de A, A^{-1}-existe, entonces: $|A| = \frac{1}{|A^{-1}|}$

f. Si A es una matriz cuadrada de orden 2-invertible, entonces:

$$A^{-1} = \frac{1}{|A|}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \quad , \quad \text{con} \quad A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

Definición 1.21 (Matriz Inversa)**.** Sea A_n una matriz cuadrada de orden n. Si existe una matriz cuadrada B_n tal que $A_nB_n = B_nA_n = I_n$, entonces se dice que B es la inversa de A y se denota por: $B = A^{-1}$, ie.

$$AA^{-1} = A^{-1}A = I_n$$

Propiedades de la Inversa:

La inversa de A, ei. A^{-1} cumple que:

a. $(A^t)^{-1} = (A^{-1})^t$.

b. $(AB)^{-1} = B^{-1}A^{-1}$.

c. $(A^{-1})^{-1} = A$.

Definición 1.22 (Matriz Ortogonal)**.** Una matriz cuadrada A_n de orden n, se dice que es ortogonal si sus columnas como vectores son perpendiculares, ie:

$$A^tA = AA^t = I_n,$$

es decir, si $A^t = A^{-1}$. (Ortonormal si además son unitarios).

Propiedades de una matriz Ortogonal:

a. $|A| = \pm 1$.

b. El producto de un número finito de matrices ortogonales es ortogonal.

c. La Inversa y por tanto la transpuesta de una matriz ortogonal es ortogonal.

d. Dada una matriz A y una matriz *ortogonal* P, entonces:
$$|A| = |P^t A P|$$

Definición 1.23 (Vectores Linealmente-Independientes (LI))**.** Un conjunto de k-vectores, $\underline{\mathbf{x}}_1, \underline{\mathbf{x}}_2, ..., \underline{\mathbf{x}}_k$, se dice que son LI si la ecuación:
$$c_1\underline{\mathbf{x}}_1 + c_2\underline{\mathbf{x}}_2 + ... + c_k\underline{\mathbf{x}}_k = 0,$$

tiene como única solución la dada por: $c_1 = c_2 = ... = c_k = 0$, para $c_i \in \mathbb{R}$, $i = 1, 2, ..., k$

Definición 1.24 (Valores y Vectores propios de una Matriz)**.** Sea A una matriz cuadrada de orden n. Se dice que λ es un valor propio de A, asociado al vector propio $\underline{\mathbf{x}}$, si se cumple que: $A\underline{\mathbf{x}} = \lambda\underline{\mathbf{x}}$

Para hallar los valores propios de una matriz A se resuelve la siguiente ecuación característica: $|A - \lambda I| = 0$.

Al dividir el vector propio $\underline{\mathbf{x}}$ por su norma $\|\underline{\mathbf{x}}\|$, se obtiene un vector unitario, el cual se denotará por: $\underline{e} = \frac{\underline{\mathbf{x}}}{\|\underline{\mathbf{x}}\|} = \frac{\underline{\mathbf{x}}}{\sqrt{\langle \underline{\mathbf{x}}\ ,\ \underline{\mathbf{x}} \rangle}}$

Alguna propiedades sobre los valores propios de una matriz:

Sean A y B matrices cuadradas de orden n y $c \in \mathbb{R}$, entonces:

a. Una matriz A tiene al menos un valor-propio igual a cero sii A es singular (no tiene inversa), ie. sii el $|A| = 0$.

b. Si A es una matriz simétrica con valores propios reales, entonces los vectores propios asociados a valores propios diferentes son ortogonales.

c. **Teorema de Descomposición Espectral:**

Cualquier matriz simétrica A, se puede escribir como:
$$A = P \quad P^t = P^t \quad P$$

en donde: -es una matriz diagonal formada por los valores propios de A, y P-es una matriz ortogonal cuyas columnas son los vectores propios unitarios asociados a los elementos de la diagonal de , ie.
$$PP^t = P^t P = I_n.$$

d. Si A es una matriz simétrica, entonces el rango de A, $r(A)$, es igual al número de sus valores propios no nulos.

e. Si $\lambda_1, \lambda_2, \ldots, \lambda_n$, son los valores propios de A, entonces:
$$tr(A) = \sum_{i=1}^{n} \lambda_i \ , \quad \text{y} \quad |A| = \prod_{i=1}^{n} \lambda_i$$

f. Si λ es una valor propio de A, entonces λ^k es un valor propio de A^k.

g. Las matrices A y A^t tienen el mismo conjunto de valores propios, pero un vector propio de A no-necesariamente es un vector propio de A^t.

h. Si A es idempotente, entonces sus valores propios son cero o uno.

Definición 1.25 (Traza de una Matriz)**.** La traza de una matriz cuadrada A de orden n, se define como la suma de los elementos de su diagonal, ie.
$$tr(A) = \sum_{i=1}^{n} a_{ii}.$$

Alguna propiedades de la traza de una matriz:

Sean A y B matrices cuadradas de orden n y $c \in \mathbb{R}$, entonces:

a. $tr(cA) = c\ tr(A)$.

b. $tr(A \pm B) = tr(A) \pm tr(B)$.

c. $tr(AB) = tr(BA)$.

d. $tr(A^t) = tr(A)$.

e. Si B^{-1}-existe, entonces: $tr(B^{-1}AB) = tr(A)$.

f. $tr(ABC) = tr(CAB) = tr(BCA)$.

g. $tr(AA^t) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2$.

Definición 1.26 (Descomposición Espectral de una Matriz)**.** Sea A una matriz simétrica de orden n, entonces:
$$A = \lambda_1 \underline{e}_1 \underline{e}_1^t + \lambda_2 \underline{e}_2 \underline{e}_2^t + \ldots + \lambda_n \underline{e}_n \underline{e}_n^t = \sum_{i=1}^{n} \lambda_i \underline{e}_i \underline{e}_i^t$$

en donde, los $\underline{e}_i$-son los vectores propios normalizados de A-asociados a los valores propios λ_i, $i = 1, 2, \ldots, n$.

A la descomposición anterior de A-se le llama descomposición espectral de A.

Definición 1.27 (Matriz Definida y Semi-Definida Positiva)**.** Sea A una matriz simétrica de orden n. Se deice que A es semi-definida positiva si para todo vector $\underline{\mathbf{x}} \in \mathbf{R}^n$ distinto del vector nulo se cumple que:
$$\underline{\mathbf{x}}^t A \underline{\mathbf{x}} \geq 0$$

Se dice que A es Definida positiva si para todo vector $\underline{\mathbf{x}} \in \mathbb{R}^n$ distinto del vector nulo se cumple que: $\underline{\mathbf{x}}^t A \underline{\mathbf{x}} > 0$

Propiedades de una Matriz Definida Positiva:

a. Si A es una matriz DP, entonces todos sus valores propios son positivos, lo que equivale a decir que su determinante es distinto de cero y por lo tanto dicha matriz tiene inversa.

b. Sea A una matriz DP. Como A-es invertible y se tiene que por la descomposición espectral:
$$A = P \ P^t = \sum_{i=1}^{n} \lambda_i \underline{e}_i \underline{e}_i^t, \quad \text{entonces:}$$
$$A^{-1} = P \ ^{-1} P^t = \sum_{i=1}^{n} \frac{1}{\lambda_i} \underline{e}_i \underline{e}_i^t$$

Definición 1.28 (Matriz Raíz-Cuadrada)**.** La raíz-cuadrada de A es:
$$A^{1/2} = P \ ^{1/2} P^t = \sum_{i=1}^{n} \sqrt{\lambda_i} \underline{e}_i \underline{e}_i^t$$

Propiedades de la Matriz Raíz-Cuadrada:

La matriz raíz-cuadrada dada por:
$$A^{1/2} = P \ ^{1/2} P^t = \sum_{i=1}^{n} \sqrt{\lambda_i} \underline{e}_i \underline{e}_i^t$$

tiene las siguientes propiedades:

a.
$$(A^{1/2})' = A^{1/2}, \quad \text{es decir:} \ \ A^{1/2} - \text{es simétrica.}$$

b.
$$A^{1/2} A^{1/2} = A.$$

c.
$$(A^{1/2})^{-1} = P \ ^{-1/2} P^t = \sum_{i=1}^{n} \frac{1}{\sqrt{\lambda_i}} \underline{e}_i \underline{e}_i^t$$

d.
$$A^{1/2} A^{-1/2} = A^{-1/2} A^{1/2} = I, \quad \text{y} \quad A^{-1/2} A^{-1/2} = A^{-1},$$

donde $A^{-1/2} = (A^{1/2})^{-1}$

Además:

$$^{-1/2} - \text{Matriz diagonal con:} \quad 1/\sqrt{\lambda_i} \quad \text{en la diagonal.}$$

$$^{1/2} - \text{Matriz diagonal con:} \quad \sqrt{\lambda_i} \quad \text{en la diagonal.}$$

$$^{-1} - \text{Matriz diagonal con:} \quad 1/ \ _{\mathbf{i}} \quad \text{en la diagonal.}$$

Definición 1.29 (Forma Cuadrática)**.** Sea A_p una matriz simétrica de orden p y $\underline{\mathbf{x}}$ un vector $p \times 1$, entonces a las función:

$$Q(\underline{\mathbf{x}}) = \underline{\mathbf{x}}^t A \underline{\mathbf{x}}$$

se le llama forma cuadrática de $\underline{\mathbf{x}}$.

$Q(\underline{\mathbf{x}})$-es un escalar y puede expresarse alternativamente por la ecuación:

$$Q(\underline{\mathbf{x}}) = \sum_{i=1}^{p} \sum_{j=1}^{p} a_{ij} x_i x_j,$$

con a_{ij}-elementos de la matriz A y x_i, x_j elementos del vector $\underline{\mathbf{x}}$.

Si $Q(\underline{\mathbf{x}}) \geq 0$, $\forall \underline{\mathbf{x}} \neq \underline{\mathbf{0}}$, entonces A-se llama semi-definida positiva y se escribe $A \geq 0$ y si $Q(\underline{\mathbf{x}}) > 0$, $\forall \underline{\mathbf{x}} \neq \underline{\mathbf{0}}$, entonces A-se llama definida positiva y se escribe $A > 0$.

Algunas Propiedades de Formas-Cuadráticas:

a. Si $A > 0$ entonces, todos sus valores propios son positivos.

b. Si $A \geq 0$ entonces, $\lambda_i \geq 0$ para $i = 1, 2, \ldots, p$ y $\lambda_i = 0$ para algún i.

c. Si $A > 0$ entonces, A-es no-singular y en consecuencia $|A| > 0$.

d. Si $A > 0$ entonces, $A^{-1} > 0$.

e. Si $A > 0$ y C_p-es una matriz no-singular entonces, $C^t A C > 0$.

f. Sea A es una matriz simétrica de orden p y sea $\underline{\mathbf{x}}$-un vector, entonces:
$$\underline{\mathbf{x}}^t A \underline{\mathbf{x}} = t(\underline{\mathbf{x}}^t A \underline{\mathbf{x}}) = t(A \underline{\mathbf{x}} \underline{\mathbf{x}}^t)$$

1.2 Diferenciación con Vectores y Matrices

Definición 1.30 (Función Vectorial)**.** Sea f-una función que asigna a un vector $\underline{x} \in \mathbb{R}^p$ un número real, ie.

$$\begin{aligned} f \quad &: \quad \mathbb{R}^p \longrightarrow \mathbb{R} \\ \underline{x} = (x_1, x_2, \ldots, x_p) &\longrightarrow f(\underline{x}). \end{aligned}$$

La derivada de $f(\underline{x})$ con respecto a $\underline{x}$ se define como el vector dado por:

$$\frac{\partial f(\underline{x})}{\partial \underline{x}} = \begin{bmatrix} \frac{\partial f(\underline{x})}{\partial \underline{x}_1} \\ \frac{\partial f(\underline{x})}{\partial \underline{x}_2} \\ \vdots \\ \frac{\partial f(\underline{x})}{\partial \underline{x}_p} \end{bmatrix} \in \mathbb{R}^p.$$

Al vector anterior se le llama: *Vector-Gradiente*.

1.2.1 Vector-Gradiente para diferentes definiciones de $f(\underline{x})$:

1. Si $f(\underline{x}) = \underline{a}^t\underline{x} = \underline{x}^t\underline{a}$, entonces:
$$\frac{\partial f(\underline{x})}{\partial \underline{x}} = \frac{\partial(\underline{a}^t\underline{x})}{\partial \underline{x}} = \frac{\partial(\underline{x}^t\underline{a})}{\partial \underline{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_p} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \underline{a}.$$

2. Si $f(\underline{x}) = \underline{x}^t A$ (vector fila), con $A_{p\times p}$- entonces:
$$\frac{\partial f(\underline{x})}{\partial \underline{x}} = \frac{\partial(\underline{x}^t A)}{\partial \underline{x}} = A, \quad \text{A-como columnas.}$$

   Si $f(\underline{x}) = A^t\underline{x}$ (vector columna), con $A_{p\times p}$- entonces:
$$\frac{\partial f(\underline{x})}{\partial \underline{x}} = \frac{\partial(A^t\underline{x})}{\partial \underline{x}} = A, \quad \text{A-como filas.}$$

3. Derivada de una forma-cuadrática:

Si $f(\underline{x}) = \underline{x}^t A\underline{x}$, con $A_{p\times p}$-matriz entonces:

a.) $\frac{\partial f(\underline{x})}{\partial \underline{x}} = \frac{\partial(\underline{x}^t A\underline{x})}{\partial \underline{x}} = A^t\underline{x} + A\underline{x}$.

b.) Si A- es simétrica entonces: $\frac{\partial f(\underline{x})}{\partial \underline{x}} = \frac{\partial(\underline{x}^t A\underline{x})}{\partial \underline{x}} = 2A\underline{x}$.

4. Derivada de la Inversa de una Matriz: $f(\mathbf{X}) = \mathbf{X}^{-1}$

La derivada de la inversa de una matriz no-singular $\mathbf{X}_{p\times p}$ respecto a su elemento en la posición ij, ie. respecto a x_{ij} es:

a.) $\frac{\partial f(\mathbf{X})}{\partial x_{ij}} = \frac{\partial(\mathbf{X}^{-1})}{\partial x_{ij}} = -\mathbf{X}^{-1}.\frac{\partial \mathbf{X}}{\partial x_{ij}}.\mathbf{X}^{-1} = -\mathbf{X}^{-1}.\Delta_{ij}.\mathbf{X}^{-1}$, si todos los elemntos de $\mathbf{X}$-son diferentes.

b.) Si $\mathbf{X}$-es simétrica entonces:

$\frac{\partial f(\mathbf{X})}{\partial x_{ij}} = \frac{\partial(\mathbf{X}^{-1})}{\partial x_{ij}} = -\mathbf{X}^{-1}.(\Delta_{ij} + \Delta_{ji}).\mathbf{X}^{-1}$, si $i \neq j$.

$\frac{\partial f(\mathbf{X})}{\partial x_{ij}} = \frac{\partial(\mathbf{X}^{-1})}{\partial x_{ij}} = -\mathbf{X}^{-1}.\Delta_{ii}.\mathbf{X}^{-1}$, si $i = j$.

Donde, $\Delta_{ij} = \frac{\partial(\mathbf{X})}{\partial x_{ij}}$-es una matriz tal que en el lugar donde se ubica el elemento x_{ij}-tiene un uno y en los demás lugares tiene ceros.

5. Derivada del Determinante de una Matriz: $f(\mathbf{X}) = |\mathbf{X}|$

La derivada del determinante de una matriz no-singular $\mathbf{X}_{p\times p}$ respecto a su elemento en la posición ij, ie. respecto a x_{ij} es:

- $\frac{\partial f(\mathbf{X})}{\partial x_{ij}} = \frac{\partial |\mathbf{X}|}{\partial x_{ij}} = \mathbf{X}_{ij}$, en donde: $\mathbf{X}_{ij}$-es el cofactor de x_{ij}.
- La matriz de derivadas es: $\frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} = (\mathbf{X}_{ij})$.
- Para matrices simétricas $\mathbf{X}$, la matriz de derivadas es:
$$\frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} = 2\text{adj}(\mathbf{X}) - \text{Diag}[\text{adj}(\mathbf{X})] = |\mathbf{X}|[2\mathbf{X}^{-1} - \text{Diag}(\mathbf{X}^{-1})].$$
donde, Diag[adj($\mathbf{X}$)-es la matriz diagonal de la adjunta de $\mathbf{X}$.

La adjunta de $\mathbf{X}$, adj($\mathbf{X}$)-es la transpuesta de la matriz de cofactores de $\mathbf{X}$.

La matriz de cofactores de $\mathbf{X}$, se obtiene al reemplazar los elementos x_{ij} de $\mathbf{X}$ por los respectivos cofactores.

El cofactor $C_{ij} = \mathbf{X}_{ij}$ de $\mathbf{X}$ se define como:
$$C_{iji} = (-1)^{i+j}|X_{ij}|,$$

donde, X_{ij}- es el menor del elemento x_{ij}, ie. matriz obtenida al borrar la fila i y la columna j de $\mathbf{X}$.

6. Derivada de la traza de una Matriz: $f(\mathbf{X}) = tr(\mathbf{X})$

La derivada de la traza de una matriz $\mathbf{X}_{p\times p}$ es:
$$\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = \frac{\partial [tr(\mathbf{X})]}{\partial \mathbf{X}} = I_p,$$

de donde:

- $\frac{\partial [tr(\mathbf{X}A)]}{\partial \mathbf{X}} = A'$, si $\mathbf{X}$-es no-simétrica.
- $\frac{\partial [tr(\mathbf{X}A)]}{\partial \mathbf{X}} = A + A' - \text{Diag}(A)$, si $\mathbf{X}$-es simétrica.

7. Derivada del logaritmo del determinante de una matriz:

$f(\mathbf{X}) = \text{Log}(|\mathbf{X}|)$

La derivada del logaritmo del determinante de una matriz $\mathbf{X}_{p\times p}$ es:
$$\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = \frac{\partial \text{Log}(|\mathbf{X}|)}{\partial \mathbf{X}} = \frac{1}{|\mathbf{X}|}\frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} = 2\mathbf{X}^{-1} - \text{Diag}(\mathbf{X}^{-1}).$$

Teorema 1.1 (Teorema de Maximización)**.** *Dada una matríz simétrica definida positiva* $\mathbf{B}_{p\times p}$ *y un* $b > 0$*, entonces:*
$$\frac{1}{|\ |^b} e^{-\frac{1}{2}tr(\ ^{-1}\mathbf{B})} \leq \frac{1}{|\mathbf{B}|^b}(2b)^{(pb)}e^{-bp}\ , \quad \forall \ \ \textit{matriz} \quad \textit{Def} +,$$

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

1.3 Organización y Presentación de Datos

1.3.1 Resumenes Descriptivos

Matriz de Datos

La matriz de datos muestrales se representa por:

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

Se tienen n-vectores $\underline{\mathbf{x}}_i$ para $i = 1, 2, \cdots, n$ en $\mathbb{R}^p$ con media dada por: $E[\underline{\mathbf{x}}_i] = \underline{\ }_{p\times 1}$ y matriz de var-cov dada por: $Cov[\underline{\mathbf{x}}_i] = {}_{p\times p}$.

Estadísticos descriptivos

Debido a la necesidad de hacer inferencias acerca de los parámetros poblacionales de una cierta distribución poblacional multivariada, es necesario definir algunos *Estadísticos Descriptivos*.

En el caso multivariado los parámetros poblacionales de interés son el vector de medias poblacional $\underline{\ }$, combinaciones lineales del vector de medias poblacional y la matriz de var-cov poblacional .

Los respectivos estadísticos descriptivos asociados a estos parámetros poblacionales son el vector de medias muestrales $\overline{\underline{\mathbf{x}}}$, combinaciones lineales de dicho vector y la matriz de var-cov muestrales $\mathbf{S}$ o $\mathbf{S}_n$.

Algunas Notaciones

En la definción del vector de medias muestrales y matriz de var-cov muestrales, se deben de tener en cuenta las siguientes notaciones.

En el vector de medias muestrales,

$$\overline{\underline{\mathbf{x}}}_{p\times 1} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{bmatrix}$$

se tiene que:

$$\overline{X}_k = \frac{1}{n}\sum_{j=1}^{n} x_{jk}$$

para $k = 1, 2, \cdots, p$.

En la matrix de Var-Cov muestrales:

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix} = \frac{1}{n-1}\sum_{k=1}^{n}(\underline{\mathbf{x}}_k - \overline{\underline{\mathbf{x}}})(\underline{\mathbf{x}}_k - \overline{\underline{\mathbf{x}}})^T,$$

se tiene que:

$$s_{kk} = \frac{1}{n-1}\sum_{j=1}^{n}(x_{jk} - \overline{x}_k)^2, \quad k = 1, 2, \cdots, p, \quad \text{y}$$

$$s_{ik} = \frac{1}{n-1}\sum_{j=1}^{n}(x_{ji} - \overline{x}_i)(x_{jk} - \overline{x}_k), \quad i, k = 1, 2, \cdots, p,$$

que representan la varianza muestral de $X^{(k)}$ y la covarianza muestral entre $X^{(i)}$ y $X^{(k)}$, para $i, k = 1, 2, \cdots, p$.

En la matrix de Correlaciones-muestrales:

$$\mathbf{R} = \begin{bmatrix} 1 & \gamma_{12} & \cdots & \gamma_{1p} \\ \gamma_{21} & 1 & \cdots & \gamma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p1} & \gamma_{p2} & \cdots & \gamma_{pp} \end{bmatrix},$$

se tiene que:

$$\gamma_{ik} = \frac{s_{ik}}{\sqrt{s_{ii}}\sqrt{s_{kk}}}, \quad i, k = 1, 2, \cdots, p,$$

que representan la correlación muestral entre $X^{(i)}$ y $X^{(k)}$, para $i, k = 1, 2, \cdots, p$.

γ_{ik}-no depende de las unidades de medida.

$\gamma_{ik} = \gamma_{ki}$.

$-1 \leq \gamma_{ik} \leq 1$.

γ_{ik}-es la covarianza muestral de las observaciones estandarizadas.

1.3.2 Representación Gráfica de Observaciones multivariadas

Algunas gráficas importantes son:

- Histogramas
- Diagramas de barras
- Bix-Plot
- Gráficos de dispersión y matriz de dispersión
- Gráficos tridimensionales y de contornos
- Gráficos de estrellas
- Gráfico de caras de Chernoff
- Gráficos de normales bivariadas
- Gráficos de Cluster

1.3.3 Vectores y Matrices Aleatorias

Definición 1.31 (Vector Aleatorio)**.** Un vector aleatorio es aquel cuyas componentes son variables aleatorias.

Definición 1.32 (Matriz Aleatoria)**.** Una matriz aleatoria es aquella cuyas componentes son variables aleatorias.

Definición 1.33 (Valor esperado de una matriz (o vector) Aleatoria)**.** El valor esperado de una matriz aleatoria (o de un vector aleatorio) es una matriz (o un vector) cuyos elementos son los valores esperados de cada entrada de la matriz, ie. el valor esperado de cada una de las variables aleatorias consideradas.

Definición 1.34 (Vector de medias poblacional)**.** Dado un vector-aleatorio:

$$\underline{\mathbf{x}} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$$

el vector de medias está dado por:

$$\underline{\mu} = E[\underline{\mathbf{x}}] = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_p] \end{bmatrix}$$

Definición 1.35 (Matriz de Var-Cov poblacional)**.** Dado un vector-aleatorio

$$\underline{\mathbf{x}} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$$

la matriz de Var-Cov de $\underline{\mathbf{x}}$ está dado por:

$$:= E[(\underline{\mathbf{x}} - \underline{\ })(\underline{\mathbf{x}} - \underline{\ })']$$

$$= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix} = \begin{bmatrix} Var[X_1] & Cov[X_1, X_2] & \cdots & Cov[X_1, X_p] \\ Cov[X_2, X_1] & Var[X_2] & \cdots & Cov[X_2, X_p] \\ \vdots & \vdots & \ddots & \vdots \\ Cov[X_p, X_1] & Cov[X_p, X_2] & \cdots & Var[X_p] \end{bmatrix}$$

donde,

$$\sigma_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)] = Cov(X_i, X_j) \quad y \quad \sigma_{ii} = E[(X_i - \mu_i)^2] = Var(X_i)$$

Definición 1.36 (Media Marginal de Xi)**.**

$$\mu_i = E[X_i] = \begin{cases} \int_{\mathbb{R}} x_i f_i(x_i) dx_i \;\; ; \;\; \text{para} \;\; X_i - \text{contínua} \\ \\ \sum x_i p_i(x_i) \;\; ; \;\; \text{para} \;\; X_i - \text{discreta} \end{cases}$$

A μ_i-se le llama la media poblacional marginal de X_i.

Definición 1.37 (Varianza Marginal de Xi)**.**

$$\sigma_i^2 = Var[X_i] = \begin{cases} \int_{\mathbb{R}} (x_i - \mu_i)^2 f_i(x_i) dx_i \;\; ; \;\; \text{para} \;\; X_i - \text{cont.} \\ \\ \sum (x_i - \mu_i)^2 p_i(x_i) \;\; ; \;\; \text{para} \;\; X_i - \text{discreta} \end{cases}$$

A σ_i^2-se le llama la varianza poblacional marginal de X_i.

$\Rightarrow$ El comportamiento conjunto de cada par de variables aleatorias X_i y X_k está descrito por su función de distribución **conjunta**.

Definición 1.38 (La covarianza poblacional)**.** Es una medida de la asociación lineal en la población de las variables X_i y X_k:

$$\sigma_{ik} = E[(X_i - \mu_i)(X_k - \mu_k)] \quad , \quad \text{donde:}$$

$$Cov[X_i, X_k] = \sigma_{ik} = \begin{cases} \int_{\mathbb{R}}\int_{\mathbb{R}}(x_i - \mu_i)(x_k - \mu_k)f_{ik}(x_i, x_k)dx_i dx_k \\ \quad \text{para} \;\; X_i, X_k - \text{contínuas} \\ \\ \sum\sum(x_i - \mu_i)(x_k - \mu_k)p_{ik}(x_i, x_k) \\ \quad \text{para} \;\; X_i, X_k - \text{discretas} \end{cases}$$

A σ_{ik}-se le llama la covarianza poblacional de X_i y X_k.

Definición 1.39 (Matriz de Correlación poblacional)**.** Dado un vector-aleatorio

$$\underline{\mathbf{x}} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$$

la matriz de Correlación de $\underline{\mathbf{x}}$ está dado por:

$$\rho = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & \rho_{pp} \end{bmatrix} = \begin{bmatrix} Corr[X_1, X_1] & Corr[X_1, X_2] & \cdots & Corr[X_1, X_p] \\ Corr[X_2, X_1] & Corr[X_2, X_1] & \cdots & Corr[X_2, X_p] \\ \vdots & \vdots & \ddots & \vdots \\ Corr[X_p, X_1] & Corr[X_p, X_2] & \cdots & Corr[X_p, X_p] \end{bmatrix}$$

donde,

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}} = Corr(X_i, X_j)$$

Relación entre y :

Dada la matriz diagonal:

$$^{1/2} :== \begin{bmatrix} \sqrt{\sigma_{11}} & & & \\ & \sqrt{\sigma_{22}} & & \\ & & \ddots & \\ & & & \sqrt{\sigma_{pp}} \end{bmatrix}$$

se cumple que:

$$= \;^{1/2}\rho\;^{1/2}$$

y

$$\rho = (\;^{1/2})^{-1}\;(\;^{1/2})^{-1}$$

Ejemplo 1.1 (Distribución Bi-Variada Discreta)**.** Hallar la matriz de var-cov del vector aleatorio bi-dimensional $\underline{\mathbf{x}}^t = (X_1, X_2)$, cuya función de densidad de probabilidad conjunta está dada por:

		X_2		
		0	1	$P_1(X_1)$
	−1	0.24	0.06	0.3
X_1	0	0.16	0.14	0.3
	1	0.4	0.0	0.4
	$P_2(X_2)$	0.8	0.2	1

Primero se halla: $\mu_1 = E[X_1] = 0.1, \quad \mu_2 = E[X_2] = 0.2,$

$$\sigma_{11} = E[(X_1 - \mu_1)^2] = \sum_{x_1} (x_1 - \mu_1)^2 P_1(x_1) = 0.69$$

y

$$\sigma_{22} = E[(X_2 - \mu_2)^2] = \sum_{x_2} (x_2 - \mu_2)^2 P_2(x_2) = 0.16$$

$$\sigma_{12} = E[(X_1 - \mu_1)(X_2 - \mu_2)] = \sum_{x_1, x_2} (x_1 - \mu_1)(x_2 - \mu_2) P_{12}(x_1, x_2) = -0.08,$$

luego:

$$\underline{\mu} = E[\underline{\mathbf{x}}] = \begin{bmatrix} E[X_1] \\ E[X_2] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix} \quad \text{y}$$

$$= E[(\underline{\mathbf{x}} - \underline{\mu})(\underline{\mathbf{x}} - \underline{\mu})^t]$$

$$= E \begin{bmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 \end{bmatrix}$$

$$= \begin{bmatrix} E(X_1 - \mu_1)^2 & E(X_1 - \mu_1)(X_2 - \mu_2) \\ E(X_2 - \mu_2)(X_1 - \mu_1) & E(X_2 - \mu_2)^2 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} 0.69 & -0.08 \\ -0.08 & 0.16 \end{bmatrix}$$

Luego,

$$^{1/2} :== \begin{bmatrix} \sqrt{\sigma_{11}} & \\ & \sqrt{\sigma_{22}} \end{bmatrix} = \begin{bmatrix} \sqrt{0.69} & \\ & \sqrt{0.16} \end{bmatrix}$$

de donde:

$$(\ ^{1/2})^{-1} := \begin{bmatrix} \frac{1}{\sqrt{\sigma_{11}}} & \\ & \frac{1}{\sqrt{\sigma_{22}}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{0.69}} & \\ & \frac{1}{\sqrt{0.16}} \end{bmatrix}$$

y por lo tanto: $\rho = (\ ^{1/2})^{-1} \Sigma (\ ^{1/2})^{-1}$

$$= \begin{bmatrix} \frac{1}{\sqrt{0.69}} & \\ & \frac{1}{\sqrt{0.16}} \end{bmatrix} \begin{bmatrix} 0.69 & -0.08 \\ -0.08 & 0.16 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{\sigma_{11}}} & \\ & \frac{1}{\sqrt{\sigma_{22}}} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -0.24 \\ -0.24 & 1 \end{bmatrix}$$

$$\rho = \begin{bmatrix} 1 & Corr(X_1, X_2) \\ Corr(X_2, X_1) & 1 \end{bmatrix}$$

Ejemplo 1.2 (Distribución Bi-Variada Discreta-2)**.** Suponga que la matriz de Var-Cov del vector-aleatorio tri-dimensional:
$\underline{\mathbf{x}} = (X_1, X_2, X_3)^t$ es:

$$= \begin{bmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$

Hallar: ρ.

1.4 Vectores y Matrices Poblacionales Particionados

1.4.1 Particionamiento del Vector de Medias-Poblacionales

Sea $\underline{\mathbf{x}}$-un vector-aleatorio p-dimensional, entonces se tienen los siguientes particionamientos para el vector de variables aleatoria $\underline{\mathbf{x}}$, el vector de medias poblacionales _ y la matriz de var-cov poblacionales , dadas por:

$$\underline{\mathbf{x}} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} X_1 \\ \vdots \\ X_q \\ \cdots \\ X_{q+1} \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} \underline{\mathbf{x}}^{(1)} \\ \cdots \\ \underline{\mathbf{x}}^{(2)} \end{bmatrix} \quad \text{y} \quad \underline{\mu} = E[\underline{\mathbf{x}}] = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_q \\ \cdots \\ \mu_{q+1} \\ \vdots \\ \mu_p \end{bmatrix} = \begin{bmatrix} \underline{\mu}^{(1)} \\ \cdots \\ \underline{\mu}^{(2)} \end{bmatrix}$$

1.4.2 Particionamiento de la Matriz de Var-Cov Poblacional

$$= \begin{array}{c} \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1q} & | & \sigma_{1,q+1} & \cdots & \sigma_{1p} \\ \vdots & & \vdots & | & \vdots & & \vdots \\ \sigma_{q1} & \cdots & \sigma_{qq} & | & \sigma_{q,q+1} & \cdots & \sigma_{qp} \\ -- & -- & -- & -- & -- & -- & -- \\ \sigma_{q+1,1} & \cdots & \sigma_{q+1,q} & | & \sigma_{q+1,q+1} & \cdots & \sigma_{q+1,p} \\ \vdots & & \vdots & | & \vdots & & \vdots \\ \sigma_{p,1} & \cdots & \sigma_{p,q} & | & \sigma_{p,q+1} & \cdots & \sigma_{p,p} \end{bmatrix} \\ q \qquad p-q \end{array}$$

$$= \begin{array}{c} q \\ \\ p-q \end{array} \begin{bmatrix} _{11} & | & _{12} \\ -- & -- & -- \\ _{21} & | & _{22} \end{bmatrix}$$

La anterior matriz particionada se obtiene como sigue:

$$(\underline{\mathbf{x}}^{(1)} - \underline{\mu}^{(1)})(\underline{\mathbf{x}}^{(2)} - \underline{\mu}^{(2)})^t = \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_q - \mu_q \end{bmatrix} \begin{bmatrix} X_{q+1} - \mu_{q+1} & \cdots & X_p - \mu_p \end{bmatrix}$$

$$= \begin{bmatrix} (X_1 - \mu_1)(X_{q+1} - \mu_{q+1}) & \cdots & (X_1 - \mu_1)(X_p - \mu_p) \\ (X_2 - \mu_2)(X_{q+1} - \mu_{q+1}) & \cdots & (X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & & \vdots \\ (X_q - \mu_q)(X_{q+1} - \mu_{q+1}) & \cdots & (X_q - \mu_q)(X_p - \mu_p) \end{bmatrix}$$

Tomando el valor esperado a ambos lados se tiene:

$$E[(\underline{\mathbf{x}}^{(1)} - \underline{\mu}^{(1)})(\underline{\mathbf{x}}^{(2)} - \underline{\mu}^{(2)})^t]$$

$$= \begin{bmatrix} E[(X_1-\mu_1)(X_{q+1}-\mu_{q+1})] & \cdots & E[(X_1-\mu_1)(X_p-\mu_p)] \\ E[(X_2-\mu_2)(X_{q+1}-\mu_{q+1})] & \cdots & E[(X_2-\mu_2)(X_p-\mu_p)] \\ \vdots & & \vdots \\ E[(X_q-\mu_q)(X_{q+1}-\mu_{q+1})] & \cdots & E[(X_q-\mu_q)(X_p-\mu_p)] \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{1,q+1} & \sigma_{1,q+2} & \cdots & \sigma_{1p} \\ \vdots & \vdots & & \vdots \\ \sigma_{q,q+1} & \sigma_{q,q+2} & \cdots & \sigma_{qp} \end{bmatrix} = {}_{12}$$

Similarmente para ${}_{11}$, ${}_{22}$ y ${}_{21} = {}^{t}_{12}$.

La matriz particionada completa se obtiene multiplicando los siguientes vectores particionados:

$$[(\underline{\mathbf{x}} - \underline{\mu})(\underline{\mathbf{x}} - \underline{\mu})^t]$$

$$= \left[\begin{pmatrix} \underline{\mathbf{x}}^{(1)} \\ \cdots \\ \underline{\mathbf{x}}^{(2)} \end{pmatrix} - \begin{pmatrix} \underline{\mu}^{(1)} \\ \cdots \\ \underline{\mu}^{(2)} \end{pmatrix} \right] \left[\begin{pmatrix} \underline{\mathbf{x}}^{(1)} \\ \cdots \\ \underline{\mathbf{x}}^{(2)} \end{pmatrix} - \begin{pmatrix} \underline{\mu}^{(1)} \\ \cdots \\ \underline{\mu}^{(2)} \end{pmatrix} \right]^t$$

$$= \begin{bmatrix} (\underline{\mathbf{x}}^{(1)} - \underline{\mu}^{(1)}) \\ (\underline{\mathbf{x}}^{(2)} - \underline{\mu}^{(2)}) \end{bmatrix} \left[(\underline{\mathbf{x}}^{(1)} - \underline{\mu}^{(1)})^t \ \ \vdots \ \ (\underline{\mathbf{x}}^{(2)} - \underline{\mu}^{(2)})^t \right]$$

$$= \begin{bmatrix} (\underline{\mathbf{x}}^{(1)} - \underline{\mu}^{(1)})(\underline{\mathbf{x}}^{(1)} - \underline{\mu}^{(1)})^t & (\underline{\mathbf{x}}^{(1)} - \underline{\mu}^{(1)})(\underline{\mathbf{x}}^{(2)} - \underline{\mu}^{(2)})^t \\ (\underline{\mathbf{x}}^{(2)} - \underline{\mu}^{(2)})(\underline{\mathbf{x}}^{(1)} - \underline{\mu}^{(1)})^t & (\underline{\mathbf{x}}^{(2)} - \underline{\mu}^{(2)})(\underline{\mathbf{x}}^{(2)} - \underline{\mu}^{(2)})^t \end{bmatrix}, \text{ luego}$$

$$= E[(\underline{\mathbf{x}} - \underline{\mu})(\underline{\mathbf{x}} - \underline{\mu})^t] = \begin{bmatrix} {}_{11} & {}_{12} \\ {}_{21} & {}_{22} \end{bmatrix}$$

1.5 Propiedades Sobre la Media y Varianza de Combinaciones Lineales

Sean X_1 y X_2 variables aleatorias y $a, b, c \in \mathbb{R}$, entonces:

1. $E[cX_1] = cE[X_1]$
2. $E[aX_1 + bX_2] = E[X_1] + bE[X_2] = a\mu_1 + b\mu_2$.

Notar que si, $aX_1 + bX_2 = (a \;\; b)\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \underline{\mathbf{c}}^t\underline{\mathbf{x}}$, de donde:

$$E[aX_1 + bX_2] = E[\underline{\mathbf{c}}^t\underline{\mathbf{x}}] = a\mu_1 + b\mu_2 = (a \;\; b)\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \underline{\mathbf{c}}^t\underline{\mu}$$

con $\underline{\mathbf{c}} = (a \;\; b)^t$ y $\underline{\mu} = (\mu \;\; \mu_2)^t$.

3. $Var[cX_1] = c^2 Var[X_1]$
4. $Cov[aX_1, bX_2] = abCov[X_1, X_2] = ab\sigma_{12}$
5.
$$\begin{aligned} Var[aX_1 + bX_2] &= a^2 Var[X_1] + b^2 Var[X_2] + 2abCov[X_1, X_2] \\ &= a^2\sigma_{11} + b^2\sigma_{22} + 2ab\sigma_{12} \end{aligned}$$

   Si, $= \begin{bmatrix} _{11} & _{12} \\ _{21} & _{22} \end{bmatrix}$, entonces:

$$\begin{aligned} \underline{\mathbf{c}}^t\Sigma\underline{\mathbf{c}} &= [a \quad b]\begin{bmatrix} _{11} & _{12} \\ _{21} & _{22} \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} \\ &= a^2\sigma_{11} + b^2\sigma_{22} + 2ab\sigma_{12} \\ &= Var(\underline{\mathbf{c}}^t\underline{\mathbf{x}}) = Var[aX_1 + bX_2] \end{aligned}$$

   En resumen:
$$Var[aX_1 + bX_2] = Var(\underline{\mathbf{c}}^t\underline{\mathbf{x}}) = \underline{\mathbf{c}}^t \;\; \underline{\mathbf{c}}$$

   En general, dado vector aleatorio $\underline{\mathbf{x}} = (X_1, X_2, \cdots, X_p)^t$, con matriz de Var-Cov dada por $_{p\times p}$, si $\underline{\mathbf{c}} = (c_1, c_2, \cdots, c_p)^t$ es un vector de constantes, entonces:
$$\begin{aligned} E[\underline{\mathbf{c}}^t\underline{\mathbf{x}}] &= E[c_1X_1 + c_2X_2 + \cdots + c_pX_p] \\ &= \underline{\mathbf{c}}^t \underline{\;}, \end{aligned}$$

   y
$$\begin{aligned} Var(\underline{\mathbf{c}}^t\underline{\mathbf{x}}) &= Var[c_1X_1 + c_2X_2 + \cdots + c_pX_p] \\ &= \underline{\mathbf{c}}^t \;\; \underline{\mathbf{c}}. \end{aligned}$$

6. Sea $\mathbf{C} = [(c_{ij})]_{q\times p}$ una matriz de constantes y $\underline{\mathbf{x}} = (X_1, X_2, \cdots, X_p)^t$ un vector aleatorio con vector de medias $\underline{\mu}$ y matriz de Var-Cov $_{p\times p}$.

Para el vector aleatorio $\underline{\mathbf{z}}_{q\times 1}$ definido como: $\underline{\mathbf{z}} = \mathbf{C}\underline{\mathbf{x}}$, se tiene que:
$$E[\underline{\mathbf{z}}] = E[\mathbf{C}\underline{\mathbf{x}}] = \mathbf{C}\underline{\mu}$$

y
$$Cov[\underline{\mathbf{z}}] = Cov[\mathbf{C}\underline{\mathbf{x}}] = \mathbf{C} \;\; \mathbf{C}^t$$

Ejemplo 1.3 (Propiedades de Medias)**.** Sea $\underline{\mathbf{x}} = (X_1 \;\; X_2)^t$-un vector aleatorio con media $\underline{\mu} = (\mu_1 \;\; \mu_2)$ y matriz de Var-Cov dada por: $= \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$.

Sea el vector aleatorio $\underline{\mathbf{z}}_{2\times 1} = (Z_1 \;\; Z_2)^t$ cuyas componentes están dadas por: $Z_1 = X_1 - X_2$ y $Z_2 = X_1 + X_2$, calcular la media y la matriz de Var-Cov de $\underline{\mathbf{z}}$.

Solución:

El vector $\underline{\mathbf{z}}$ se puede escribir como sigue:

$$\underline{\mathbf{z}} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \mathbf{C}\underline{\mathbf{x}}$$

luego, usando el resultado anterior se tiene que:

$$E[\underline{\mathbf{z}}] = E[\mathbf{C}\underline{\mathbf{x}}] = \mathbf{C}\underline{\mu} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_1 + \mu_2 \end{bmatrix}$$

$$\begin{aligned} Cov[\underline{\mathbf{z}}] &= E[\mathbf{C}\underline{\mathbf{x}}] = \mathbf{C}\ \mathbf{C}^t \\ &= \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{11} - \sigma_{12} & \sigma_{12} - \sigma_{22} \\ \sigma_{11} + \sigma_{12} & \sigma_{12} + \sigma_{22} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{11} - \sigma_{12} - \sigma_{12} + \sigma_{22} & \sigma_{11} - \sigma_{12} + \sigma_{12} - \sigma_{22} \\ \sigma_{11} + \sigma_{12} - \sigma_{12} - \sigma_{22} & \sigma_{11} + \sigma_{12} + \sigma_{12} + \sigma_{22} \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{11} - 2\sigma_{12} + \sigma_{22} & \sigma_{11} - \sigma_{22} \\ \sigma_{11} - \sigma_{22} & \sigma_{11} + 2\sigma_{12} + \sigma_{22} \end{bmatrix} \end{aligned}$$

$$Cov[\underline{\mathbf{z}}] = \begin{bmatrix} Var[Z_1] & Cov[Z_1, Z_2] \\ Cov[Z_2, Z_1] & Var[Z_2] \end{bmatrix}$$

1.6 Vectores y Matrices Muestrales Particionadas

1.6.1 Particionamiento del Vector de Medias Muestrales

Sea $\underline{\mathbf{x}}$-un vector-aleatorio p-dimensional, entonces se tienen los siguientes particionamientos para el vector de variables aleatoria $\underline{\mathbf{x}}$, el vector de medias muestrales $\overline{\mathbf{x}}$ y la matriz de var-cov muestrales $\mathbf{S}$, dadas por:

$$\underline{\mathbf{x}} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} X_1 \\ \vdots \\ X_q \\ \cdots \\ X_{q+1} \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} \underline{\mathbf{x}}^{(1)} \\ \cdots \\ \underline{\mathbf{x}}^{(2)} \end{bmatrix} \quad \text{y} \quad \overline{\underline{\mathbf{x}}} = \begin{bmatrix} \overline{X}_1 \\ \vdots \\ \overline{X}_q \\ \cdots \\ \overline{X}_{q+1} \\ \vdots \\ \overline{X}_p \end{bmatrix} = \begin{bmatrix} \overline{\underline{\mathbf{x}}}^{(1)} \\ \cdots \\ \overline{\underline{\mathbf{x}}}^{(2)} \end{bmatrix}$$

1.6.2 Particionamiento de la Matriz de Var-Cov Muestrales

$$\mathbf{S} = \left[\begin{array}{ccc|ccc}
s_{11} & \cdots & s_{1q} & s_{1,q+1} & \cdots & s_{1p} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
s_{q1} & \cdots & s_{qq} & s_{q,q+1} & \cdots & s_{qp} \\
\hline
s_{q+1,1} & \cdots & s_{q+1,q} & s_{q+1,q+1} & \cdots & s_{q+1,p} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
s_{p,1} & \cdots & s_{p,q} & s_{p,q+1} & \cdots & s_{p,p}
\end{array}\right]$$

$$= \begin{array}{cc}
 & \begin{array}{cc} q & p-q \end{array} \\
\begin{array}{c} q \\ p-q \end{array} & \left[\begin{array}{c|c} S_{11} & S_{12} \\ \hline S_{21} & S_{22} \end{array}\right]
\end{array}$$

1.7 Algunas formas matriciales Eficientes

Para la matriz de datos

$$\mathbf{X}_{n\times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \quad \mathbf{X}^T = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{bmatrix}$$

y

$$\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n\times 1}$$

Se tienen las siguientes expresiones para ciertas estadísticas de resúmenes:

1. $\underline{\bar{\mathbf{x}}} = \begin{bmatrix} \overline{X}_1 \\ \overline{X}_2 \\ \vdots \\ \overline{X}_p \end{bmatrix} = \frac{1}{n}\mathbf{X}^T\mathbf{1} = \frac{1}{n}\begin{bmatrix} \sum_{i=1}^n x_{i1} \\ \sum_{i=1}^n x_{i2} \\ \vdots \\ \sum_{i=1}^n x_{ip} \end{bmatrix}_{p\times 1}$

2. $\mathbf{S} = \frac{1}{n-1}\mathbf{X}^T[\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T]\mathbf{X} = \frac{1}{n-1}\mathbf{X}^T\mathbf{H}\mathbf{X}$, con $\mathbf{H} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$

3. $\mathbf{R} = D^{-\frac{1}{2}}\mathbf{S}D^{-\frac{1}{2}}$, en donde $D^{-\frac{1}{2}}$, es una matriz diagonal cuyos elementos son los inversos de las desviaciones estándar muestrales, es decir que:

$$D^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{\sqrt{s_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{s_{22}}} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{s_{pp}}} \end{bmatrix}$$

También se cumple que: $\mathbf{S} = D^{\frac{1}{2}}\mathbf{R}D^{\frac{1}{2}}$

4. Se define la Varianza-Generalizada de $\mathbf{X}$ como el determinante de $\mathbf{S}$,
$$VG = |\mathbf{S}|,$$
la cual representa una medida de variabilidad del vector de variables aleatorias $\underline{\mathbf{x}}$

5. Se define la Varianza Total de $\mathbf{X}$ como la traza de $\mathbf{S}$, ie
$$VT = Tr(\mathbf{S})$$

1.8 Muestra Aleatoria de Distribuciones p-Variadas

Sea un vector p-variado de variables aleatorias: $\underline{\mathbf{x}} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$, con función de distribución multivariada representada por $f(\underline{\mathbf{x}})$, con media $\underline{\ }$ y var-cov , ie, $E[\underline{\mathbf{x}}] = \underline{\ }$ y $\mathrm{Var}(\underline{\mathbf{x}}) =$.

Una muestra aleatoria de tamaño n de esta distribución es un conjunto de n-vectores aleatorios p-variados, $\underline{\mathbf{x}}_1, \underline{\mathbf{x}}_2, \cdots, \underline{\mathbf{x}}_n$, independientes e identicamente distribuídos con distribución $f(\underline{\mathbf{x}})$, es decir que:
$$h(\underline{\mathbf{x}}_1, \underline{\mathbf{x}}_2, \cdots, \underline{\mathbf{x}}_n) = h_1(\underline{\mathbf{x}}_1)h_2(\underline{\mathbf{x}}_2)\ldots h_n(\underline{\mathbf{x}}_n) = \prod_{i=1}^{n} f(\underline{\mathbf{x}}_i)$$

con $E[\underline{\mathbf{x}}_i] = \underline{\ }$, $\mathrm{Var}(\underline{\mathbf{x}}_i) =$ y $Cov(\underline{\mathbf{x}}_j, \underline{\mathbf{x}}_k) = 0 \ \forall \ j \neq k$.

Teorema 1.2 (Propiedades de)**.** *Dada una m.a de tamaña n, $\underline{\mathbf{x}}_1, \underline{\mathbf{x}}_2, \cdots, \underline{\mathbf{x}}_n$ de una distribución p-variada, con vector de media $\underline{\ }$ y matriz de Var-Cov , ie. $E[\underline{\mathbf{x}}_i] = \underline{\mu}$ y $Var[\underline{\mathbf{x}}_i] =$, se cumple lo siguiente:*

1. $E[\overline{\underline{\mathbf{x}}}] = \underline{\ }$
2. $Var[\overline{\underline{\mathbf{x}}}] = \frac{1}{n}$
3. $E[\mathbf{S}] =$ y $E[\mathbf{S}_n] = \frac{n-1}{n}$

Proof. **Para la parte (i)** observe que:
$$\overline{\underline{\mathbf{x}}}_{p\times 1} = \begin{bmatrix} \overline{X}_1 \\ \overline{X}_2 \\ \vdots \\ \overline{X}_p \end{bmatrix} = \frac{1}{n}\begin{bmatrix} \sum_{i=1}^n x_{i1} \\ \sum_{i=1}^n x_{i2} \\ \vdots \\ \sum_{i=1}^n x_{ip} \end{bmatrix} = \frac{1}{n}\begin{bmatrix} x_{11} + x_{21} + \cdots + x_{n1} \\ x_{12} + x_{22} + \cdots + x_{n2} \\ \vdots \\ x_{1p} + x_{2p} + \cdots + x_{np} \end{bmatrix}_{p\times 1}$$
$$= 1/n[\underline{\mathbf{x}}_1 + \underline{\mathbf{x}}_2 + \cdots + \underline{\mathbf{x}}_n]$$

Luego de lo anterior se tiene que:

$$
\begin{aligned}
E[\underline{\overline{\mathbf{x}}}] &= \frac{1}{n}E[\underline{\mathbf{x}}_1 + \underline{\mathbf{x}}_2 + \cdots + \underline{\mathbf{x}}_n] \\
&= \frac{1}{n}(E[\underline{\mathbf{x}}_1] + E[\underline{\mathbf{x}}_2] + \cdots + E[\underline{\mathbf{x}}_n]) \\
&= \frac{1}{n}(\underline{\ } + \underline{\ } + \cdots + \underline{\ }) \\
&= \frac{1}{n}n\underline{\ } \\
E[\underline{\overline{\mathbf{x}}}] &= \underline{\ }
\end{aligned}
$$

Para la parte (ii), oberve que:

$$
Var(\underline{\overline{\mathbf{x}}}) = E\left[(\underline{\overline{\mathbf{x}}} - E[\underline{\overline{\mathbf{x}}}])(\underline{\overline{\mathbf{x}}} - E[\underline{\overline{\mathbf{x}}}])^T\right] = E\left[(\underline{\overline{\mathbf{x}}} - \underline{\ })(\underline{\overline{\mathbf{x}}} - \underline{\ })^T\right]
$$

$$
= E\left[\left(\frac{1}{n}\sum_{j=1}^{n}(\underline{\mathbf{x}}_j - \underline{\ })\right)\left(\frac{1}{n}\sum_{k=1}^{n}(\underline{\mathbf{x}}_k - \underline{\ })\right)^T\right]
$$

$$
= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}E(\underline{\mathbf{x}}_j - \underline{\ })(\underline{\mathbf{x}}_k - \underline{\ })^T
$$

$$
= \frac{1}{n^2}\left[\sum_{j=1}^{n}E(\underline{\mathbf{x}}_j - \underline{\ })(\underline{\mathbf{x}}_j - \underline{\ })^T + \sum_{j,k,j\neq k}E(\underline{\mathbf{x}}_j - \underline{\ })(\underline{\mathbf{x}}_k - \underline{\ })^T\right]
$$

$$
= \frac{1}{n^2}\left[\sum_{j=1}^{n}\mathrm{Var}(\underline{\mathbf{x}}_j) + \sum_{j,k,j\neq k}\mathrm{Cov}(\underline{\mathbf{x}}_j, \underline{\mathbf{x}}_k)\right]
$$

$$
= \frac{1}{n^2}[\ + \ + \cdots + \ + 0 + 0 + \cdots + 0] = \frac{1}{n^2}n\ = \frac{1}{n}
$$

Para la parte (iii), observe que:

$$
\begin{aligned}
&\sum_{j=1}^{n}(\underline{\mathbf{x}}_j - \underline{\overline{\mathbf{x}}})(\underline{\mathbf{x}}_j - \underline{\overline{\mathbf{x}}})^T = \\
&= \sum_{j=1}^{n}(\underline{\mathbf{x}}_j - \underline{\overline{\mathbf{x}}})\underline{\mathbf{x}}_j^T - \sum_{j=1}^{n}(\underline{\mathbf{x}}_j - \underline{\overline{\mathbf{x}}})\underline{\overline{\mathbf{x}}}^T \\
&= \sum_{j=1}^{n}\underline{\mathbf{x}}_j\underline{\mathbf{x}}_j^T - n\underline{\overline{\mathbf{x}}}\ \underline{\overline{\mathbf{x}}}^T,
\end{aligned}
$$

lo anterior debido a que se cumple lo siguiente:

$$
\sum_{j=1}^{n}(\underline{\mathbf{x}}_j - \underline{\overline{\mathbf{x}}}) = 0 \quad \text{y} \quad \sum_{j=1}^{n}\underline{\mathbf{x}}_j^T = n\underline{\overline{\mathbf{x}}}^T
$$

Tomando valor esperado en lo anterior se tiene que:

$$
\begin{aligned}
E\left[\sum_{j=1}^{n}(\underline{\mathbf{x}}_j-\overline{\underline{\mathbf{x}}})(\underline{\mathbf{x}}_j-\overline{\underline{\mathbf{x}}})^T\right] &= E\left[\sum_{j=1}^{n}\underline{\mathbf{x}}_j\underline{\mathbf{x}}_j^T-n\overline{\underline{\mathbf{x}}}\ \overline{\underline{\mathbf{x}}}^T\right]\\
&=\sum_{j=1}^{n}E[\underline{\mathbf{x}}_j\underline{\mathbf{x}}_j^T]-nE[\overline{\underline{\mathbf{x}}}\ \overline{\underline{\mathbf{x}}}^T],
\end{aligned}
$$

pero por propiedades del valor esperado y de la matriz de Var-Cov de un vector aleatorio $\underline{\mathbf{x}}$ se tiene que:

$$Var(\underline{\mathbf{x}}) = \quad = E[\underline{\mathbf{x}}\underline{\mathbf{x}}^T]-E[\underline{\mathbf{x}}]E[\underline{\mathbf{x}}]^T,$$

es decir que,

$$E[\underline{\mathbf{x}}\underline{\mathbf{x}}^T] = \quad + E[\underline{\mathbf{x}}]E[\underline{\mathbf{x}}]^T,$$

de donde:

$$
\begin{aligned}
&\sum_{j=1}^{n}E[\underline{\mathbf{x}}_j\underline{\mathbf{x}}_j^T]-nE[\overline{\underline{\mathbf{x}}}\ \overline{\underline{\mathbf{x}}}^T]\\
&=\sum_{j=1}^{n}\left[Var(\underline{\mathbf{x}}_j)+E[\underline{\mathbf{x}}_j]E[\underline{\mathbf{x}}_j]^T\right]-n\left[Var(\overline{\underline{\mathbf{x}}})+E[\overline{\underline{\mathbf{x}}}]E[\overline{\underline{\mathbf{x}}}]^T\right]\\
&=\sum_{j=1}^{n}(\ +\underline{\ }\ \underline{\ }^{\,T})-n\left(\frac{1}{n}\ +\underline{\ }\ \underline{\ }^{\,T}\right)\\
&=n\ -n\underline{\ }\ \underline{\ }^{\,T}-\ -n\underline{\ }\ \underline{\ }^{\,T}\\
&=(n-1)\ ,
\end{aligned}
$$

es decir que,

$$E\left[\sum_{j=1}^{n}(\underline{\mathbf{x}}_j-\overline{\underline{\mathbf{x}}})(\underline{\mathbf{x}}_j-\overline{\underline{\mathbf{x}}})^T\right]=(n-1)$$

Pero

$$\mathbf{S}=\frac{1}{n-1}\sum_{j=1}^{n}(\underline{\mathbf{x}}_j-\overline{\underline{\mathbf{x}}})(\underline{\mathbf{x}}_j-\overline{\underline{\mathbf{x}}})^T,$$

y por lo tanto se tiene que:

$$
\begin{aligned}
E[\mathbf{S}] &= \frac{1}{n-1}E\left[\sum_{j=1}^{n}(\underline{\mathbf{x}}_j-\overline{\underline{\mathbf{x}}})(\underline{\mathbf{x}}_j-\overline{\underline{\mathbf{x}}})^T\right]\\
&=\frac{1}{n-1}(n-1)\\
E[\mathbf{S}] &=
\end{aligned}
$$

$\square$

Resumen: EL teorema anterior dice que el vector de medias muestrales $\overline{\underline{\mathbf{x}}}$ es un estimador insesgado del vector de medias poblacionales $\underline{\ }$ y que la matriz de Var-Cov muestrales $\mathbf{S}$ también es un estimador insesgado de la matriz de Var-Cov poblacionales , pero $\mathbf{S}_n$ es sesgado.

1.9 Distancias

1.9.1 Introducción

Muchas técnicas importantes del análisis multivariado se basan en el concepto de distancia.

Al medir distancias entre variables, se obtiene una idea de la proximidad entre ellas.\

Definición 1.40 (Distancia Euclídea). Dados dos vectores $\underline{\mathbf{x}}, \underline{\mathbf{y}} \in \mathbb{R}^p$, $\underline{\mathbf{x}} = (x_1, x_2, \dots, x_p)$ y $\underline{\mathbf{y}} = (y_1, y_2, \dots, y_p)$, la distancia euclídea entre ellos se define como sigue:

$$d(\underline{\mathbf{x}}, \underline{\mathbf{y}}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2} = \|\underline{\mathbf{x}} - \underline{\mathbf{y}}\|$$

Desde el punto de vista estadístico, la distancia euclídea no es muy-satisfatoria, ya que cada coordenada está ponderada por un mismo factor de 1.

Cuando las coordenadas representan medidas sujetas a fluctuaciones aleatorias de diferentes magnitudes (por ejemplo: altura en metros, peso en kilogramos, distancia en kilómetros, etc), es preferible ponderar las coordenadas de acuerdo a su variabilidad.

Es usual usar ponderaciones pequeñas para las coordenadas sujetas a un alto grado de variabilidad.

Debido a esto es necesario desarrollar una distancia que tenga en cuenta la variabilidad y la dependencia (correlación) entre las variables.

Definición 1.41 (Distancia Estadística). Dado un vector aleatorio p-dimensional $\underline{\mathbf{x}} = (X_1, X_2, \dots, X_p)$ cuyas componentes X_i´s son independientes, se define la distancias estadística de $\underline{\mathbf{x}}$ al origen de coordenadas $\underline{\mathbf{0}} = (0, 0, \dots, 0)$ en $\mathbb{R}^p$ como sigue:

$$d(\underline{\mathbf{0}}, \underline{\mathbf{x}}) = \sqrt{\frac{x_1^2}{s_{11}} + \frac{x_2^2}{s_{22}} + \dots + \frac{x_p^2}{s_{pp}}}$$

Similarmente dados dos vectores aleatorios $\underline{\mathbf{x}} = (X_1, X_2, \dots, X_p)$ y $\underline{\mathbf{y}} = (Y_1, Y_2, \dots, Y_p)$ en $\mathbb{R}^p$ de la misma población, entonces la distancia estadística entre $\underline{\mathbf{x}}$ y $\underline{\mathbf{y}}$ está dada por:

$$d(\underline{\mathbf{x}}, \underline{\mathbf{y}}) = \sqrt{\frac{(x_1 - y_1)^2}{s_{11}} + \frac{(x_2 - y_2)^2}{s_{22}} + \dots + \frac{(x_p - y_p)^2}{s_{pp}}}$$

Si $s_{11} = s_{22} = \cdots = s_{pp}$, entonces se puede usar la distancia euclídea con pesos-iguales.

Si las variables de los vectores anteriores no son independientes, entonces las expresiones anteriores no son adecuadas, por lo que hay que definir otras distancias que tengan en cuenta la correlación (o covarianzas) entre las variables.

Definición 1.42 (Distancia de Mahalanobis)**.** La distancia de Mahalanobis entre dos vectores aleatorios $\underline{\mathbf{x}} = (X_1, X_2, \ldots, X_p)$ y $\underline{\mathbf{y}} = (Y_1, Y_2, \ldots, Y_p)$ en $\mathbb{R}^p$ se define como sigue:

$$d(\underline{\mathbf{x}}, \underline{\mathbf{y}}) = (\underline{\mathbf{x}} - \underline{\mathbf{y}})^t \ {}^{-1}(\underline{\mathbf{x}} - \underline{\mathbf{y}}), \quad \text{si} \qquad \text{es conocida}$$

y por

$$d(\underline{\mathbf{x}}, \underline{\mathbf{y}}) = (\underline{\mathbf{x}} - \underline{\mathbf{y}})^t S^{-1}(\underline{\mathbf{x}} - \underline{\mathbf{y}}), \quad \text{si} \qquad \text{es desconocida}$$

La distancia de Mahalanobis entre un vector aleatorio $\underline{\mathbf{x}} = (X_1, X_2, \ldots, X_p)$ y su vector de medias $\_ = (\mu_1, \mu_2, \ldots, \mu_p)$ en $\mathbb{R}^p$ se define como sigue:

$$d(\underline{\mathbf{x}}, \_) = (\underline{\mathbf{x}} - \_)^t \ {}^{-1}(\underline{\mathbf{x}} - \_), \quad \text{si} \quad \text{y} \quad \underline{\mu} \quad \text{son conocidas}$$

y por

$$d(\underline{\mathbf{x}}, \overline{\mathbf{x}}) = (\underline{\mathbf{x}} - \overline{\mathbf{x}})^t S^{-1}(\underline{\mathbf{x}} - \overline{\mathbf{x}}), \quad \text{si} \quad \text{y} \quad \underline{\mu} \quad \text{son desconocidas}$$

Ejemplo 1.4. Los Biólogos **Grojan** y **Wirth** (1981) descubrieron dos nuevas especies de insectos **Ameroheleafasciata** (AF) y **Apseudofasciata** (APF). Puesto que las especies son similares en apariencia, resulta útil para el biólogo estar en capacidad de clasificar un especimen como AF o APF basado en características externas que son fáciles de medir. Entre alguna de las características que distinguen los AF de los APF, los biólogos reportan medidas de la longitud de las antenas (X) y la longitud de las alas (Y), ambas medidas en milímetros, de nueve insectos AF y de seis insectos APF.

Una de las preguntas que motivaron el estudio fue: ¿Será posible encontrar una regla que nos permita clasificar un insecto dado como AF o como APF, basados únicamente en las mediciones de las antenas y las alas? Rta: Si.

Los datos son los siguientes:

ESPECIE	X	Y
AF	1.38	1.64
AF	1.40	1.20
AF	1.24	1.72
AF	1.36	1.74
AF	1.38	1.82
AF	1.48	1.82
AF	1.54	1.82
AF	1.38	1.90
AF	1.56	2.08
APF	1.14	1.78
APF	1.20	1.86
APF	1.18	1.96
APF	1.30	1.96
APF	1.26	2.00
APF	1.28	2.00

Para el conjunto de datos dado, responda las siguientes preguntas:

1. Construya un gráfico de X e Y. Comente acerca de la apariencia de los datos.

2. Para cada grupo de individuos, calcule el vector de medias muestrales, la matriz de varianzas-covarianzas muestrales, la matriz de correlaciones muestrales, la varianza total y la varianza generalizada. Realice la descomposición espectral de cada una de las dos matrices de Var-Cov asociadas a cada grupo de individuos.

3. Calcule la distancia euclídea entre los vectores de media de cada grupo de individuos.

4. Calcule la distancia de Mahalanobis entre los vectores de media de cada grupo de individuos.

5. ¿Considera razonable usar la distancia de Mahalanobis en cada uno de los dos grupos?

Chapter 2

Distribución Normal Multivariada

2.1 Geometría y propiedades de la NM

Existen varias formas de identificar si un vector aleatorio es distribuido normal multivariado, o mejor dicho de identificar si es o no es un vector aleatorio con distribución de probabilidad normal-multivariado, entre ellas están:

- Usando la f.d.p. de dicho vector.
- Usando una caracterización de la distribución NM basada en la normal univariada.
- Usando la función generadora de momentos.

2.2 Normal Univariada

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

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

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

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

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

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

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

2.3 Normal Multivariada

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

$$(\underline{\mathbf{x}} - \_)^t\ ^{-1}(\underline{\mathbf{x}} - \_),$$

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

Definición 2.1 (Normal Multivariada)**.** Sea $\underline{\mathbf{x}}$ un vector aleatorio p-dimensional, ie. $\underline{\mathbf{x}} \in \mathbb{R}^p$. Se dice que $\underline{\mathbf{x}}$ tiene una distribución aleatoria Normal-Multivariada (o normal p-variada), con vector de medias $\underline{\mu}$ y matriz de Var-Cov , lo cual se denota por: $\underline{\mathbf{x}} \sim N_p(\underline{\mu},\ )$, si su f.d.p multivariada está dada por:

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

2.3.1 Algunos Aspectos Geométricos de la NM

- La expresión $(\underline{\mathbf{x}} - \_)^t\ ^{-1}(\underline{\mathbf{x}} - \_) = c^2$, 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: $\pm c\sqrt{\lambda_i}\underline{\mathbf{e}}_i$ para $i = 1, 2, \ldots, p$, donde.los λ_i-son los valores propios de asociados a los respectivos vectores propios $\underline{\mathbf{e}}_i$.

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

Teorema 2.1. *Si -es definida-positiva y λ es un valor propio de asociado al vector propio $\underline{\mathbf{e}}$ (ie. $\ \underline{\mathbf{e}} = \lambda\underline{\mathbf{e}}$), entonces $\frac{1}{\lambda}$-es un valor-propio de $\ ^{-1}$-asociado al mismo vector propio $\underline{\mathbf{e}}$ (ie. $\ ^{-1}\underline{\mathbf{e}} = \frac{1}{\lambda}\underline{\mathbf{e}}$). Además, $\ ^{-1}$-también es definida positiva.*

Proof. Como -es definida-positiva, se cumple que: $\underline{\mathbf{e}}^t\ \underline{\mathbf{e}} > 0$, con $\underline{\mathbf{e}} \neq 0$.

Ahora, para $\underline{\mathbf{e}}$-vector-propio de se cumple lo siguiente:

$$\ \underline{\mathbf{e}} = \lambda\underline{\mathbf{e}}, \quad \text{luego,}$$

$$0 < \underline{\mathbf{e}}^t\ \underline{\mathbf{e}} = \underline{\mathbf{e}}^t\lambda\underline{\mathbf{e}} = \lambda\underline{\mathbf{e}}^t\underline{\mathbf{e}} = \lambda,$$

es decir, $\lambda > 0$.

Ahora,

$$\begin{aligned} \underline{\mathbf{e}} &= \ ^{-1}(\ \underline{\mathbf{e}}) \\ &= \ ^{-1}(\lambda\underline{\mathbf{e}})\ , \quad \text{pues} \quad \ \underline{\mathbf{e}} = \lambda\underline{\mathbf{e}} \\ \text{luego} \quad \frac{1}{\lambda}\underline{\mathbf{e}} &= \ ^{-1}\underline{\mathbf{e}}, \end{aligned}$$

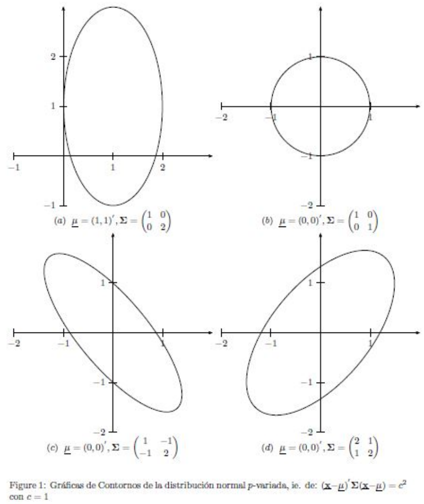

Figure 2.1: Gráfico de Contornos NM

es decir, $\frac{1}{\lambda}$-es un valor-propio de $\ ^{-1}$-asociado al mismo vector propio $\underline{\mathbf{e}}$.

Veamos que $\ ^{-1}$-es Definida positiva. (Def.+).

Para un vector $\underline{\mathbf{x}} \neq 0$, se cumple lo siguiente:

$$\begin{aligned}
\underline{\mathbf{x}}^t\ ^{-1}\underline{\mathbf{x}} &= \underline{\mathbf{x}}^t\left(\sum_{i=1}^{p}\frac{1}{\lambda_i}\underline{\mathbf{e}}_i\underline{\mathbf{e}}_i^t\right)\underline{\mathbf{x}}\ , \quad \text{por la desc. espectral de} \quad ^{-1} \\
&= \sum_{i=1}^{p}\frac{1}{\lambda_i}\underline{\mathbf{x}}^t\underline{\mathbf{e}}_i\underline{\mathbf{e}}_i^t\underline{\mathbf{x}} \\
&= \sum_{i=1}^{p}\frac{1}{\lambda_i}(\underline{\mathbf{e}}_i^t\underline{\mathbf{x}})^2\ , \ \text{pues} \ \ (\underline{\mathbf{e}}_i^t\underline{\mathbf{x}})^2 = (\underline{\mathbf{e}}_i^t\underline{\mathbf{x}})^t(\underline{\mathbf{e}}_i^t\underline{\mathbf{x}}) = \underline{\mathbf{x}}^t\underline{\mathbf{e}}_i\underline{\mathbf{e}}_i^t\underline{\mathbf{x}} \\
&> 0, \quad \text{pues los} \ \ \lambda_i > 0 \ \ \text{y} \ \ \underline{\mathbf{x}} \neq 0.
\end{aligned}$$

Luego, $\ ^{-1}$-es definida positiva. □

Ejemplo 2.1 (Aspectos Geométricos-1)**.** Suponga que un vector aleatorio 2-dimensional $\underline{\mathbf{x}}$ tiene la siguiente matriz de Var-Cov,

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

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: $\pm c\sqrt{\lambda_i}\underline{\mathbf{e}_i}$, para $i = 1, 2$ y $c > 0$, es decir: $\pm c\sqrt{\lambda_1}\underline{\mathbf{e}_1}$ y $\pm c\sqrt{\lambda_2}\underline{\mathbf{e}_2}$.

Ahora se deben hallar los valores propios de , para lo cual se debe resolver la ecuación característica dada por:

$$| \ - \lambda I_2| = 0.$$

$$\begin{aligned} | \ - \lambda I_2| &= \left| \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{11} \end{pmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right| \\ &= \begin{vmatrix} \sigma_{11} - \lambda & \sigma_{12} \\ \sigma_{12} & \sigma_{11} - \lambda \end{vmatrix} \\ &= (\sigma_{11} - \lambda)^2 - \sigma_{12}^2 \\ &= (\sigma_{11} - \lambda - \sigma_{12})(\sigma_{11} - \lambda + \sigma_{12}) \end{aligned}$$

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

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

Ahora los vectores propios asociados a estos valores propios se obtienen resolviendo la ecuación: $\underline{\mathbf{e}} = \lambda \underline{\mathbf{e}}$.

Por Ejemplo si, $\underline{\mathbf{e}}_1 = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$-es el vector propio asociado al valor propio $\lambda_1 = \sigma_{11} + \sigma_{12}$, entonces debe cumplir que:

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{11} \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = (\sigma_{11} + \sigma_{12}) \begin{pmatrix} e_1 \\ e_2 \end{pmatrix},$$

de donde, se obtiene el siguiente sistema de ecuaciones:

$$\begin{aligned} \sigma_{11} e_1 + \sigma_{12} e_2 &= (\sigma_{11} + \sigma_{12}) e_1 \implies e_1 = e_2 \\ \sigma_{12} e_1 + \sigma_{11} e_2 &= (\sigma_{11} + \sigma_{12}) e_2 \implies e_1 = e_2 \end{aligned}$$

El vector, $\underline{\mathbf{e}}_1 = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} e_1 \\ e_1 \end{bmatrix}$-normalizado está dado por:

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

De manera análoga, se encuentra que el segundo vector-propio asociado al segundo valor-propio $\lambda_2 = \sigma_{11} - \sigma_{12}$ es:

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

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: $\lambda_1 = \sigma_{11} + \sigma_{12} > \sigma_{11} > \sigma_{11} - \sigma_{12} = \lambda_2$.

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° y que pasa por el punto $\underline{\mu} = (\mu_1, \mu_2)$.

Ejemplo 2.2 (Aspectos Geométricos-2)**.** Suponga que el vector aleatorio $\underline{\mathbf{x}}^t = (X_1, X_2)$ tiene una distribución normal-bivariada, con vector de medias $\underline{\mu} = (\mu_1, \mu_2)$ y matriz de Var-Cov dada por . Halla la f.d.p multivariada de $\underline{\mathbf{x}}$.

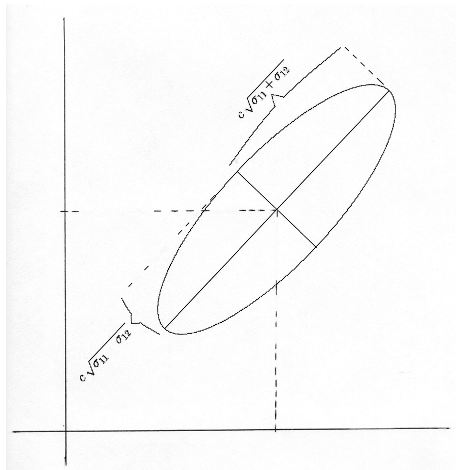

Figure 2.2: Gráfico de Elipse NB

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

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

Para $\ = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$

se tiene que:

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

con $|\ | = \sigma_{11}\sigma_{22} - \sigma_{12}^2$, pero $\sigma_{12} = \rho\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}$,

es decir que, $|\ | = \sigma_{11}\sigma_{22} - \rho^2\sigma_{11}\sigma_{22} = \sigma_{11}\sigma_{22}(1-\rho^2)$, de donde:

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

Ahora se tiene que:

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

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

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

de donde,

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

y como $|\ | = \sigma_{11}\sigma_{22}(1-\rho^2)$, entonces $|\ |^{-\frac{1}{2}} = \left[\sigma_{11}\sigma_{22}(1-\rho^2)\right]^{-\frac{1}{2}}$

por lo tanto,

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

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

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

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

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

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

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

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

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

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

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

En el gráfico (a) se observa la superficie (densidad) de una distribución normal-bivarianda con $\sigma_{11} = \sigma_{22}$ y $\rho_{12} = Corr(X_1, X_2) = 0$

En el gráfico (b) se observa la superficie (densidad) de una distribución normal-bivarianda con $\sigma_{11} = \sigma_{22}$ y $\rho_{12} = Corr(X_1, X_2) = 0.75$.

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 $(\underline{\mathbf{x}} - \_)^t\ ^{-1}(\underline{\mathbf{x}} - \_)$ es igual a cero, es decir, cuando $\underline{\mathbf{x}} = \_$. Por tanto el punto $\_$ es el punto de máxima densidad, o la moda, y también es la media, como se observa en la siguiente gráfica.

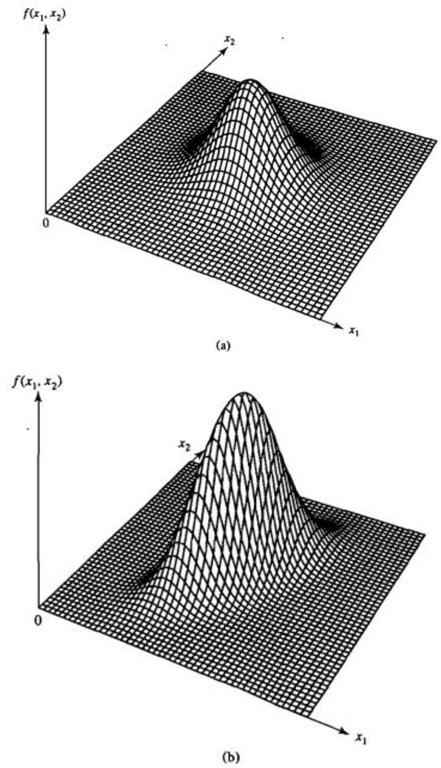

Figure 2.3: Gráfico de Superficies NB

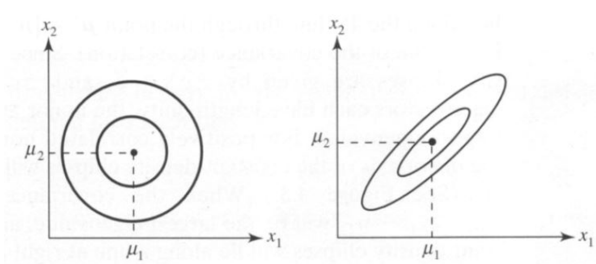

Figure 2.4: Gráfico de Contornos de Superficies NB

2.4 Propiedades de la distribución Normal Multivariada

Sea $\underline{\mathbf{x}}$-un vector aleatorio p-variado.

1. Si $\underline{X} \sim N_p(\underline{\mu},\ )$, entonces:
$$E[\underline{X}] = \underline{\mu} \quad \text{y} \quad Var[\underline{X}] = \tag{2.1}$$

La distribución de $\underline{X}$-queda completamente determinada por $\underline{\mu}$ y .

2. Si $\underline{X} \sim N_p(\underline{\mu},\ )$, entonces
$$\underline{a}^t\underline{X} = a_1X_1 + a_2X_2 + ... + a_pX_p \sim N_1(\underline{a}^t\underline{\mu}\ ,\ \underline{a}^t\ \underline{a}) \tag{2.2}$$
Análogamente, si $\forall\ \underline{a} \in \mathbb{R}^p\ \underline{a}^t\underline{X} \sim N_1(\underline{a}^t\underline{\mu}\ ,\ \underline{a}^t\ \underline{a})$ entonces $\underline{X} \sim N_p(\underline{\mu},\ )$, es decir:
$$\underline{X} \sim N_p(\underline{\mu},\ ) \Longleftrightarrow \underline{a}^t\underline{X} \sim N_1(\underline{a}^t\underline{\mu}\ ,\ \underline{a}^t\ \underline{a})$$
Luego, si $\underline{X} \sim N_p(\underline{\mu},\ )$ entonces cada $X_i \sim N_1(\mu_i\ ,\ \sigma_{ii})$, 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 dados por:
$$\_ = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \quad \text{y} \quad = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

Definición 2.2 (Función Generadora de Momentos-NU)**.** Si una v.a $X \sim N_1(\mu\ \ \sigma^2)$, entonces la FGM de X es:
$$M_X(t) := E[e^{tX}] = e^{\mu t + \frac{1}{2}\sigma^2 t^2},\ t \in \mathbb{R}$$

Definición 2.3 (Función Generadora de Momentos-NM)**.** Si un vector aleatorio $\underline{X} \sim N_p(\underline{\mu},\ )$, entonces la FGM de $\underline{X}$ es:
$$M_{\underline{X}}(\underline{t}) := E[e^{\underline{t}^t\underline{X}}] = e^{\underline{t}^t\underline{\mu} + \frac{1}{2}\underline{t}^t\ \underline{t}},\ \ \underline{t} \in \mathbb{R}^p$$

Desmostración prop.2. $\underline{X} \sim N_p(\underline{\mu},\ ) \Longleftrightarrow \underline{a}^t\underline{X} \sim N_1(\underline{a}^t\underline{\mu}\ ,\ \underline{a}^t\ \underline{a})$

Sea la v.a $Y = \underline{a}^t\underline{X}$
$$\begin{aligned} M_Y(t) &= E[e^{tY}] = E[e^{t(\underline{a}^t\underline{X})}] \\ &= E[e^{(\underline{a}t)^t\underline{X}}]\ ,\ \text{pues}\ \ (\underline{a}t)^t = t^t\underline{a}^t = t\underline{a}^t \\ &= M_{\underline{X}}(\underline{a}t) \\ &= e^{(\underline{a}t)^t\underline{\mu} + \frac{1}{2}(\underline{a}t)^t\ (\underline{a}t)} \\ &= e^{t^t\underline{a}^t\underline{\mu} + \frac{1}{2}t^t\underline{a}^t\ \underline{a}t} \\ &= e^{t(\underline{a}^t\underline{\mu}) + \frac{1}{2}t(\underline{a}^t\ \underline{a})t} = e^{t(\underline{a}^t\underline{\mu}) + \frac{1}{2}(\underline{a}^t\ \underline{a})t^2} \end{aligned}$$

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

□

Ejemplo 2.3. Suponga que $\underline{\mathbf{x}} \sim N_3(\underline{\mu}\ ,\quad)$ y considere la combinación linea de las componentes de $\underline{\mathbf{x}}$ dada por:

$$\underline{\mathbf{y}} = 2X_1 - X_2 + 3X_3 = \begin{bmatrix} 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \underline{a}^t \underline{\mathbf{x}}$$

con $\underline{a} = \begin{bmatrix} 2 & -1 & 3 \end{bmatrix}^t$, entonces:

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

$$\begin{aligned} Var[\underline{\mathbf{y}}] &= Var[\underline{a}^t \underline{\mathbf{x}}] \\ &= \underline{a}^t \quad \underline{a} \\ &= \begin{bmatrix} 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} \\ &= \begin{bmatrix} 2\sigma_{11} - \sigma_{12} + 3\sigma_{13} & 2\sigma_{12} - \sigma_{22} + 3\sigma_{23} & 2\sigma_{13} - \sigma_{23} + 3\sigma_{33} \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} \\ &= 4\sigma_{11} - 2\sigma_{12} + 6\sigma_{13} - 2\sigma_{12} + \sigma_{22} - 3\sigma_{23} + 6\sigma_{13} - 3\sigma_{23} + 9\sigma_{33} \end{aligned}$$

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

Es decir que:

$$\underline{\mathbf{y}} = \underline{a}^t \underline{X} = 2X_1 - X_2 + 3X_3 \sim N_1(\underline{a}^t \underline{\mu}\ ,\ \underline{a}^t \quad \underline{a}).$$

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

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

3. Si $\underline{\mathbf{x}} \sim N_p(\underline{\mu}\ ,\quad)$, entonces

$$\underline{\mathbf{y}} = A\underline{X} + \underline{b} \sim N_q(A\underline{\mu} + \underline{b}\ ,\ A \quad A^t) \tag{2.3}$$

Si $\underline{b} = \underline{0}$ se tiene que: $\underline{\mathbf{y}} = A\underline{X} \sim N_q(A\underline{\mu}\ ,\ A \quad A^t)$, con $A_{q\times p}$.

Proof.

$$
\begin{aligned}
M_{\underline{\mathbf{y}}}(\underline{t}) &:= E\left[\ e^{\underline{t}^t\underline{\mathbf{y}}}\right] = E\left[\ e^{\underline{t}^t(A\underline{\mathbf{x}}+\underline{b})}\right] \\
&= E\left[\ e^{\underline{t}^tA\underline{\mathbf{x}}+\underline{t}^t\underline{b}}\right] \\
&= E\left[\ e^{\underline{t}^tA\underline{\mathbf{x}}}e^{\underline{t}^t\underline{b}}\right] \\
&=\ e^{\underline{t}^t\underline{b}}E\left[\ e^{\underline{t}^tA\underline{\mathbf{x}}}\right] \\
&=\ e^{\underline{t}^t\underline{b}}E\left[\ e^{(A^t\underline{t})^t\underline{\mathbf{x}}}\right]\ ,\ \text{pues}\ \ (A^t\underline{t})^t = \underline{t}^tA \\
&=\ e^{\underline{t}^t\underline{b}}M_{\underline{X}}(A^t\underline{t}) \\
&=\ e^{\underline{t}^t\underline{b}}\left[\ e^{(A^t\underline{t})^t\underline{\mu}+\frac{1}{2}(A^t\underline{t})^t\ (A^t\underline{t})}\ \right];\ M_{\underline{X}}(\underline{t}) = e^{\underline{t}^t\underline{\mu}+\frac{1}{2}\underline{t}^t\ \underline{t}} \\
&=\ e^{\underline{t}^t\underline{b}}\left[\ e^{\underline{t}^tA\underline{\mu}+\frac{1}{2}\underline{t}^tA\ A^t\underline{t}}\ \right] \\
&=\ e^{\underline{t}^t\left(A\underline{\mu}+\underline{b}\right)+\frac{1}{2}\underline{t}^t(A\ A^t)\underline{t}^t}
\end{aligned}
$$

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

$$\underline{\mathbf{y}} = A\underline{X} + \underline{b} \sim N_q(A\underline{\mu} + \underline{b}\ ,\ A\ A^t)\ ,\quad l.q.q.d.$$

$\square$

Ejemplo 2.4. Suponga que $\underline{\mathbf{x}} \sim N_3(\underline{\mu}\ ,\quad)$ y considere el vector de combinaciones lineales de las variables de $\underline{\mathbf{x}}$ dado por:

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

entonces,

$$\underline{\mathbf{z}} = A\underline{X} \sim N_q(A\underline{\mu}\ ,\ A\ A^t)$$

donde

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

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

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

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

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

Ejemplo 2.5. Suponga que $\underline{\mathbf{x}}^T = (X_1, X_2, X_3) \sim N_3(\underline{\mu}\ ,\ \ )$, 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 $Z_2 = X_1 - X_2$.

$$\begin{aligned} \text{Sea :} \quad \underline{\mathbf{z}} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} &= \begin{bmatrix} X_1 + X_2 + X_3 \\ X_1 - X_2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \\ &= A\underline{\mathbf{x}} \end{aligned}$$

luego,

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

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

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

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.

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

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

entonces:

$$\underline{\mathbf{x}}^{(1)} \sim N_q\left(\underline{\mu}^{(1)}\ ,\ \ {}_{11}\right) \tag{2.4}$$

$$\underline{\mathbf{x}}^{(2)} \sim N_{p-q}\left(\underline{\mu}^{(2)}\ ,\ \ {}_{22}\right). \tag{2.5}$$

La anterior propiedad también se puede enunciar como sigue: Todos los subconjuntos de variables de $\underline{\mathbf{x}}$ tienen distribución normal, sea univariada o multivariada.

Proof. Para esto se usarán las siguientes matrices particionadas:

$$A_{q\times p} = \begin{array}{cccc} & q & & p-q \\ q & [\ \mathbf{I}_q & \vdots & \mathbf{O}]_{q\times p} \end{array} = [\ \mathbf{I}_q \quad \vdots \quad \mathbf{O}]_{q\times p} \quad \text{y} \quad \underline{b} = \underline{0}_{q\times 1}$$

Luego,

$$
\begin{aligned}
A\underline{X} + \underline{b} &= A\underline{X} \\
&= \begin{bmatrix} \mathbf{I}_q & \vdots & \mathbf{O} \end{bmatrix}_{q\times p} \begin{bmatrix} \underline{\mathbf{x}}^{(1)} \\ \cdots \\ \underline{\mathbf{x}}^{(2)} \end{bmatrix}_{p\times 1} \\
&= \begin{bmatrix} \mathbf{I}_q\ \underline{\mathbf{x}}^{(1)} & + & \mathbf{O}\ \underline{\mathbf{x}}^{(2)} \end{bmatrix}_{q\times 1} \\
&= \begin{bmatrix} \mathbf{I}_q\ \underline{\mathbf{x}}^{(1)} \end{bmatrix}_{q\times 1} \\
&= \begin{bmatrix} \underline{\mathbf{x}}^{(1)} \end{bmatrix}_{q\times 1}
\end{aligned}
$$

de donde, usando la propiedad (2.3) se tiene que:

$$A\underline{X} + \underline{b} = \underline{\mathbf{x}}^{(1)} \sim N_q(A\underline{\mu}\ ,\ A\ A^t)$$

Pero,

$$
\begin{aligned}
A\underline{\mu} &= \begin{bmatrix} \mathbf{I}_q & \vdots & \mathbf{O} \end{bmatrix} \begin{bmatrix} \underline{\mu}^{(1)} \\ \cdots \\ \underline{\mu}^{(2)} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{I}_q\ \underline{\mu}^{(1)} & + & \mathbf{O}\ \underline{\mu}^{(2)} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{I}_q\ \underline{\mu}^{(1)} \end{bmatrix} \\
&= \begin{bmatrix} \underline{\mu}^{(1)} \end{bmatrix}_{q\times 1}
\end{aligned}
$$

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

Es decir que:
$$\underline{\mathbf{x}}^{(1)} \sim N_q\left(A\underline{\mu}\ ,\ A\ \ A^t\right)$$
$$N_q\left(\underline{\mu}^{(1)}\ ,\ \ {}_{11}\right)\ ,\qquad l.q.q.d.$$

□

De manera similar, se demuestra que:
$$\underline{\mathbf{x}}^{(2)} \sim N_q\left(\underline{\mu}^{(2)}\ ,\ \ {}_{22}\right),$$

en este caso usando las matrices particionadas dadas por:
$$A = \begin{array}{cc} & \begin{array}{ccc} q & & p-q \end{array} \\ p-q & \left[\begin{array}{ccc} \mathbf{O} & \vdots & \mathbf{I}_{p-q} \end{array}\right]_{(p-q)\times p} \end{array} = \left[\begin{array}{ccc} \mathbf{O} & \vdots & \mathbf{I}_{p-q} \end{array}\right]\ ;\ \underline{b} = \underline{0}_{(p-q)\times 1}$$

Ejemplo 2.6. Sea $\underline{\mathbf{x}} \sim N_5(\underline{\mu}\ ,\ \ )$, hallar la distribución de: $\begin{bmatrix} X_2 \\ X_4 \end{bmatrix}$. Sean
$$\underline{\mathbf{x}}^{(1)} = \begin{bmatrix} X_2 \\ X_4 \end{bmatrix}\ ,\ \ \underline{\ }^{(1)} = \begin{bmatrix} \mu_2 \\ \mu_4 \end{bmatrix}\quad \text{y}\quad {}_{11} = \begin{bmatrix} \sigma_{22} & \sigma_{24} \\ \sigma_{24} & \sigma_{44} \end{bmatrix}$$

Asumiendo que $\underline{\mathbf{x}}$, $\underline{\ }$ y son particionados como sigue:
$$\underline{\mathbf{x}} = \begin{bmatrix} X_2 \\ X_4 \\ \cdots \\ X_1 \\ X_3 \\ X_5 \end{bmatrix}\ ,\ \ \underline{\ } = \begin{bmatrix} \mu_2 \\ \mu_4 \\ \cdots \\ \mu_1 \\ \mu_3 \\ \mu_5 \end{bmatrix}\quad \text{y}\quad = \begin{bmatrix} \sigma_{22} & \sigma_{24} & \vdots & \sigma_{12} & \sigma_{23} & \sigma_{25} \\ \sigma_{24} & \sigma_{44} & \vdots & \sigma_{14} & \sigma_{34} & \sigma_{45} \\ \cdots & \cdots\cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_{12} & \sigma_{14} & \vdots & \sigma_{11} & \sigma_{13} & \sigma_{15} \\ \sigma_{23} & \sigma_{34} & \vdots & \sigma_{13} & \sigma_{33} & \sigma_{35} \\ \sigma_{25} & \sigma_{45} & \vdots & \sigma_{15} & \sigma_{35} & \sigma_{55} \end{bmatrix}$$
$$\text{es decir :}\quad \underline{\mathbf{x}} = \begin{bmatrix} \underline{\mathbf{x}}^{(1)} \\ \cdots \\ \underline{\mathbf{x}}^{(2)} \end{bmatrix}\ ,\ \ \underline{\mu} = \begin{bmatrix} \underline{\mu}^{(1)} \\ \cdots \\ \underline{\mu}^{(2)} \end{bmatrix}\quad \text{y}\quad = \begin{bmatrix} {}_{11} & | & {}_{12} \\ -- & -- & -- \\ {}_{21} & | & {}_{22} \end{bmatrix}$$
$$\text{luego :}\quad \underline{\mathbf{x}}^{(1)} \sim N_2\left(\underline{\mu}^{(1)}\ ,\ \ {}_{11}\right) \sim N_2\left(\begin{bmatrix} \mu_2 \\ \mu_4 \end{bmatrix}\ ,\begin{bmatrix} \sigma_{22} & \sigma_{24} \\ \sigma_{24} & \sigma_{44} \end{bmatrix}\right).$$

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

entonces:
$$\underline{\mathbf{x}}^{(1)} \perp \underline{\mathbf{x}}^{(2)}\ ,\ \ \text{si y solo si}\quad {}_{12} = \ {}^{t}_{21} = O_{q\times(p-q)}. \tag{2.6}$$

es decir, $\underline{\mathbf{x}}^{(1)}$ y $\underline{\mathbf{x}}^{(2)}$ son estadísticamente independientes si y solo si: ${}_{12} = \ {}^{t}_{21} = O$.

Ejemplo 2.7. Suponga que $\underline{\mathbf{x}} \sim N_3(\underline{\mu}\ ,\ \ )$, con
$$= \begin{bmatrix} 4 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

¿Son las variables X_1 y X_2 independientes? Rta. NO, porque $\sigma_{12} = 1 \neq 0$.

¿Son los siguientes vectores aleatorios independientes?:

$\begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$ y $[X_3]$.

Sean

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

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

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

por lo tanto, si son independientes.

b). Si $\underline{\mathbf{x}}^{(1)} \perp \underline{\mathbf{x}}^{(2)}$ tales que:

$\underline{\mathbf{x}}^{(1)} \sim N_q\left(\underline{\mu}^{(1)}, {}_{11}\right)$ y $\underline{\mathbf{x}}^{(2)} \sim N_q\left(\underline{\mu}^{(2)}, {}_{22}\right)$ entonces,

$$\underline{\mathbf{x}} = \begin{bmatrix} \underline{\mathbf{x}}^{(1)} \\ \cdots \\ \underline{\mathbf{x}}^{(2)} \end{bmatrix} \sim N_p\left[\begin{pmatrix} \underline{\mu}^{(1)} \\ \cdots \\ \underline{\mu}^{(2)} \end{pmatrix}, \begin{pmatrix} {}_{11} & | & O \\ -- & -- & -- \\ O & | & {}_{22} \end{pmatrix}\right] \tag{2.7}$$

c). Si $\underline{\mathbf{x}}^{(1)} \perp \underline{\mathbf{x}}^{(2)}$,

$\underline{\mathbf{x}}^{(1)} \sim N_q\left(\underline{\mu}^{(1)}, {}_{11}\right)$ y $\underline{\mathbf{x}}^{(2)} \sim N_q\left(\underline{\mu}^{(2)}, {}_{22}\right)$, Además:

$$\underline{\mathbf{x}}^{(1)} + \underline{\mathbf{x}}^{(2)} \sim N_k\left(\underline{\mu}^{(1)} + \underline{\mu}^{(2)}, {}_{11} + {}_{22}\right), \quad \text{si} \quad q = p - q = k \tag{2.8}$$

Proof. Para esto se usarán la FGM de un vector normal-multivariado dada por:

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

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

$$\begin{aligned} M_{\underline{\mathbf{x}}^{(1)} + \underline{\mathbf{x}}^{(2)}}(\underline{t}) &:= E\left[e^{\underline{t}^t(\underline{\mathbf{x}}^{(1)} + \underline{\mathbf{x}}^{(2)})}\right] \\ &= E\left[e^{\underline{t}^t \underline{\mathbf{x}}^{(1)}} e^{\underline{t}^t \underline{\mathbf{x}}^{(2)})}\right] \\ &= E\left[e^{\underline{t}^t \underline{\mathbf{x}}^{(1)}}\right] E\left[e^{\underline{t}^t \underline{\mathbf{x}}^{(2)})}\right], \text{ pues } \quad \underline{\mathbf{x}}^{(1)} \perp \underline{\mathbf{x}}^{(2)} \\ &= M_{\underline{\mathbf{x}}^{(1)}}(\underline{t}) M_{\underline{\mathbf{x}}^{(2)}}(\underline{t}) \\ &= e^{\underline{t}^t \underline{\mu}^{(1)} + \frac{1}{2}\underline{t}^t {}_{11}\underline{t}} \; e^{\underline{t}^t \underline{\mu}^{(2)} + \frac{1}{2}\underline{t}^t {}_{22}\underline{t}} \\ &= e^{\underline{t}^t \underline{\mu}^{(1)} + \underline{t}^t \underline{\mu}^{(2)} + \frac{1}{2}\underline{t}^t {}_{11}\underline{t} + \frac{1}{2}\underline{t}^t {}_{22}\underline{t}} \\ &= e^{\underline{t}^t\left(\underline{\mu}^{(1)} + \underline{\mu}^{(2)}\right) + \frac{1}{2}\underline{t}^t({}_{11} + {}_{22})\underline{t}} \end{aligned}$$

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 ${}_{11} + {}_{22}$, es decir:

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

□

Demostración de 5(a) y 5(b); ver Ejercicio 4.14 del libro de Johnson, sixth edition [Johnson and Wichern, 2007].

6. Si $\underline{\mathbf{x}} = \begin{bmatrix} \underline{\mathbf{x}}^{(1)} \\ \cdots \\ \underline{\mathbf{x}}^{(2)} \end{bmatrix} \sim N_p \left[\begin{pmatrix} \underline{\mu}^{(1)} \\ \cdots \\ \underline{\mu}^{(2)} \end{pmatrix} , \begin{pmatrix} {}_{11} & | & {}_{12} \\ -- & -- & -- \\ {}_{21} & | & {}_{22} \end{pmatrix} \right]$, con $|\ {}_{22}| > 0$,

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

$$\underline{\mathbf{x}}^{(1)}|\underline{\mathbf{x}}^{(2)} = \underline{\mathbf{x}}^{(2)} \sim N_q \left[\underline{\mu}^{(1)} + {}_{12}{}_{22}^{-1} \left(\underline{\mathbf{x}}^{(2)} - \underline{\mu}^{(2)} \right) , \ {}_{11} - {}_{12}{}_{22}^{-1}{}_{21} \right] \tag{2.9}$$

es decir:

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

y

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

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

$$\underline{\mathbf{x}}^{(2)}|\underline{\mathbf{x}}^{(1)} = \underline{\mathbf{x}}^{(1)} \sim N_{p-q} \left[\underline{\mu}^{(2)} + {}_{21}{}_{11}^{-1} \left(\underline{\mathbf{x}}^{(1)} - \underline{\mu}^{(1)} \right) , \ {}_{22} - {}_{21}{}_{11}^{-1}{}_{12} \right] \tag{2.10}$$

Proof. Considere la matriz $p \times p$ dada por:

$$\mathbf{A} = \begin{bmatrix} \mathbf{I}_{q\times q} & | & - {}_{12}{}_{22}^{-1}{}_{q\times(p-q)} \\ ----- & -- & ----- \\ \mathbf{0}_{(p-q)\times q} & | & \mathbf{I}_{(p-q)\times(p-q)} \end{bmatrix}_{p\times p}$$

teniendo en cuenta que:

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

se tiene que:

$$\underline{Z}_{p\times 1} = \mathbf{A}(\underline{X} - \underline{\mu}) = \begin{bmatrix} \mathbf{I}_{q\times q} & | & - {}_{12}{}_{22}^{-1}{}_{q\times(p-q)} \\ ----- & -- & ----- \\ \mathbf{0}_{(p-q)\times q} & | & \mathbf{I}_{(p-q)\times(p-q)} \end{bmatrix} \begin{bmatrix} \underline{X}^{(1)} - \underline{\mu}^{(1)} \\ \\ \underline{X}^{(2)} - \underline{\mu}^{(2)} \end{bmatrix}$$

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

con

$$\underline{\mu}_{\underline{Z}} = \mathbf{A}\underline{0} = \underline{0}$$

y

$${}_{\underline{Z}} = \mathbf{A} \ \mathbf{A}^t$$

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

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

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

de lo anterior se tiene que como los vectores aleatorios:

$$\underline{X}^{(1)} - \underline{\mu}^{(1)} -\ _{12}\ _{22}^{-1} \left(\underline{X}^{(2)} - \underline{\mu}^{(2)}\right) \quad \text{y} \quad \underline{X}^{(2)} - \underline{\mu}^{(2)}$$

tienen matriz de var-cov $\mathbf{0}$, entonces son independientes, y además se cumple que:

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

y al ser

$$\underline{X}^{(1)} - \underline{\mu}^{(1)} -\ _{12}\ _{22}^{-1} \left(\underline{X}^{(2)} - \underline{\mu}^{(2)}\right) \quad \text{y} \quad \underline{X}^{(2)} - \underline{\mu}^{(2)}$$

independientes, entonces la distribución condicional de:

$$\underline{X}^{(1)} - \underline{\mu}^{(1)} -\ _{12}\ _{22}^{-1} \left(\underline{X}^{(2)} - \underline{\mu}^{(2)}\right) \quad \text{dado} \quad \underline{X}^{(2)} - \underline{\mu}^{(2)}$$

es igual a la marginal de:

$$\underline{X}^{(1)} - \underline{\mu}^{(1)} -\ _{12}\ _{22}^{-1} \left(\underline{X}^{(2)} - \underline{\mu}^{(2)}\right)$$

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

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

$\square$

Ejemplo 2.8. Distribución condicional de una Normal-Bivariada.

Sea $\underline{\mathbf{x}} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N_2 \left[\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}\ ,\ \left(\begin{array}{c|c} \sigma_{11} & \sigma_{12} \\ \hline \sigma_{21} & \sigma_{22} \end{array}\right) \right]$, con $\sigma_{22} > 0$,

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

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

pues,

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

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

y como,

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

luego:

$$\frac{\sigma_{12}}{\sigma_{22}} = \frac{\sqrt{\sigma_{11}}}{\sqrt{\sigma_{22}}}\rho_{12}$$

y

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

es decir:

$$\begin{aligned} X_1|X_2 = x_2 &\sim N_1\left(\mu_1 + \frac{\sqrt{\sigma_{11}}}{\sqrt{\sigma_{22}}}\rho_{12}(x_2-\mu_2)\ ,\ \sigma_{11}\left[1-\rho_{12}^2\right]\right) \\ &\sim N_1\left(\underbrace{\mu_1 - \frac{\sqrt{\sigma_{11}}}{\sqrt{\sigma_{22}}}\rho_{12}\mu_2}_{a_0} + \underbrace{\frac{\sqrt{\sigma_{11}}}{\sqrt{\sigma_{22}}}\rho_{12}}_{b}x_2\ ,\ \sigma_{11}\left[1-\rho_{12}^2\right]\right) \end{aligned}$$

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

de donde se observa que la media de la distribución condicional de $X_1|X_2 = x_2$, 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:

- En regresión multivariada, la media condicional

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

  es llamada la curva de regresión.

- Entonces la curva de regresión en la normal multivariada, $\underline{\mu}_{1.2} = E[\underline{X}^{(1)}|\underline{X}^{(2)}]$, se puede escribir como:

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

  Ahora sea

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

entonces, la curva de regresión se puede escribir como
$E[\underline{X}^{(1)}|\underline{X}^{(2)}] = \underline{\mu}^{(1)} + {}_{12}{}_{22}^{-1}\left(\underline{X}^{(2)} - \underline{\mu}^{(2)}\right) =$

$$\begin{bmatrix} E[X_1|X_{q+1}, X_{q+2}, \cdots, X_p] \\ \\ E[X_2|X_{q+1}, X_{q+2}, \cdots, X_p] \\ \vdots \\ \vdots \\ E[X_q|X_{q+1}, X_{q+2}, \cdots, X_p] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \\ \mu_2 \\ \vdots \\ \vdots \\ \mu_q \end{bmatrix} + \begin{bmatrix} \beta_{1,q+q} & \beta_{1,q+2} & \cdots & \beta_{1,p} \\ \\ \beta_{2,q+q} & \beta_{2,q+2} & \cdots & \beta_{2,p} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \beta_{q,q+q} & \beta_{q,q+2} & \cdots & \beta_{q,p} \end{bmatrix} \begin{bmatrix} X_{q+1} - \mu_{q+1} \\ \\ X_{q+2} - \mu_{q+2} \\ \vdots \\ \vdots \\ X_p - \mu_p \end{bmatrix}$$

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

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.

- La matriz de var-cov condicional
$$ {}_{1.2} = {}_{11} - {}_{12}{}_{22}^{-1}{}_{21}$$
es constante pues no depende de los valores de las variables condicionantes. Por tanto, la curva de regresión es homocedástica.

Ejemplo 2.9 (Distribución Condicional)**.** Suponga que $\underline{\mathbf{x}}^T = (X_1, X_2, X_3) \sim N_3(\underline{\mu}\,, \;)$, donde:

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

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

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

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

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

y matriz de var-cov dada por:

$$Var(\underline{\mathbf{x}}^{(1)}|\underline{\mathbf{x}}^{(2)}) = {}_{11} - {}_{12}{}_{22}^{-1}{}_{21}$$

Haciendo,

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

y

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

luego,

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

y

$$Var(\underline{\mathbf{x}}^{(1)}|\underline{\mathbf{x}}^{(2)}) = {}_{11} - {}_{12}\,{}_{22}^{-1}\,{}_{21} = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} \frac{1}{1} \begin{pmatrix} 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 3 \end{pmatrix}$$

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

$$\underline{\mathbf{z}} = {}^{-1/2}(\underline{X} - \underline{\mu}) \sim N_p(\underline{0}\ ,\ \mathbf{I}_p) \tag{2.11}$$

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

La Demostración de la propiedad (2.11) se puede estudiarla del libro de [Johnson and Wichern, 2007].

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

a.)

$$(\underline{X} - \underline{\mu})^t\ {}^{-1}(\underline{X} - \underline{\mu}) \sim \chi^2_p \tag{2.12}$$

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

$$\left\{\underline{\mathbf{x}}\ :\ (\underline{\mathbf{x}} - \underline{\mu})^t\ {}^{-1}(\underline{\mathbf{x}} - \underline{\mu}) \leq \chi^2_{\alpha;p}\right\} \tag{2.13}$$

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

Proof. Sea

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

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

entonces, por el resultado de la prop: (2.11) se tiene que:

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

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

Ahora, se Considera la variable:

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

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

$$= \underline{Z}^t\underline{Z} = \sum_{i=1}^{p} Z_i^2 \sim \chi^2_{(p)}$$

Observaciones:

- Por el teorema de descomposición espectral de la matriz simétrica y definida positiva

  $$= \quad {}^{t}$$

  donde es una matriz ortogonal con los vectores propios de como columnas, y es una matriz diagonal con los valores propios de en su diagonal. Entonces,

  $$= \quad {}^{t} = \quad {}^{1/2}\ {}^{1/2}\ {}^{t} = (\quad {}^{1/2}\ {}^{t})(\quad {}^{1/2}\ {}^{t}) = \quad {}^{1/2}\ {}^{1/2}$$

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

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

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

9. Sean $\underline{\mathbf{x}}_1, \underline{\mathbf{x}}_2, \dots, \underline{\mathbf{x}}_n$, vectores aleatorios p-variados mutuamente-independientes tales que:

   $$\underline{\mathbf{x}}_i \sim N_p(\underline{\mu}_i\ ,\quad)\ , \quad \text{para} \quad i = 1, 2, \dots, n,$$

   entonces:

   $$\begin{aligned} \underline{\mathbf{v}}_1 &= \sum_{i=1}^{n} c_i \underline{\mathbf{x}}_i \\ &= c_1 \underline{\mathbf{x}}_1 + c_2 \underline{\mathbf{x}}_2 + \dots + c_n \underline{\mathbf{x}}_n \\ &\sim N_p \left(\sum_{i=1}^{n} c_i \underline{\mu}_i\ ,\ \left(\sum^{n} c_i^2 \right) \quad \right) \qquad (2.14) \\ &\sim N_p \left(c_1 \underline{\mu}_1 + c_2 \underline{\mu}_2 + \dots + c_n \underline{\mu}_n\ ,\ [c_1^2 + c_2^2 + \dots + c_n^2] \quad \right) \end{aligned}$$

además, si

$$\underline{\mathbf{v}}_2 = \sum_{i=1}^{n} b_i \underline{\mathbf{x}}_i = b_1 \underline{\mathbf{x}}_1 + b_2 \underline{\mathbf{x}}_2 + \dots + b_n \underline{\mathbf{x}}_n$$

entonces:

$$\underline{\mathbf{v}} = \begin{bmatrix} \underline{\mathbf{v}}_1 \\ \cdots \\ \underline{\mathbf{v}}_2 \end{bmatrix} \sim N_{2p} \left[\begin{pmatrix} \sum_{i=1}^{n} c_{i-i} \\ \cdots\cdots \\ \sum_{i=1}^{n} b_{i-i} \end{pmatrix}\ ,\ {}_{\underline{\mathbf{v}}_1, \underline{\mathbf{v}}_2} \right]\ , \quad \text{donde} \qquad (2.15)$$

$$\Sigma_{\underline{\mathbf{v}}_1, \underline{\mathbf{v}}_2} = Cov(\underline{\mathbf{v}}_1\ ,\ \underline{\mathbf{v}}_2) = \begin{bmatrix} \left(\sum_{i=1}^{n} c_i^2\right) & \vdots & \left(\sum_{i=1}^{n} b_i c_i\right) \\ \cdots\cdots & & \cdots\cdots \\ \left(\sum_{i=1}^{n} b_i c_i\right) & \vdots & \left(\sum_{i=1}^{n} b_i^2\right) \end{bmatrix}$$

$$= \begin{bmatrix} (\underline{c}^t \underline{c}) & \vdots & \left(\underline{b}^t \underline{c}\right) \\ \cdots\cdots & & \cdots\cdots \\ \left(\underline{b}^t \underline{c}\right) & \vdots & \left(\underline{b}^t \underline{b}\right) \end{bmatrix}_{2p \times 2p}$$

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

Proof. Considere el vector de $np \times 1$ dado por:

$$\underline{Z} = \begin{bmatrix} \underline{X}_1 \\ \underline{X}_2 \\ \vdots \\ \vdots \\ \underline{X}_n \end{bmatrix}_{np\times 1}$$

por el resultado de la propiedad 5c (2.8) se tiene que:

$$\underline{Z} \sim N_{np}(\underline{\mu}_{\underline{Z}} \ , \quad _{\underline{Z}})$$

donde,

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

□

para la parte (a):

Considere la matriz

$$\mathbf{A} = [c_1\mathbf{I} \quad | \quad c_2\mathbf{I} \quad | \quad \cdots \quad | \quad c_n\mathbf{I}]_{p\times np}$$

donde $\mathbf{I}$ es la matriz identidad de orden $p \times p$, entonces

$$\mathbf{A}\underline{Z} = [c_1\mathbf{I} \quad | \quad c_2\mathbf{I} \quad | \quad \cdots \quad | \quad c_n\mathbf{I}] \begin{bmatrix} \underline{X}_1 \\ \underline{X}_2 \\ \vdots \\ \vdots \\ \underline{X}_n \end{bmatrix} = c_1\underline{X}_1 + c_2\underline{X}_2 + \cdots + c_n\underline{X}_n = \underline{V}_1$$

y por un resultado anterior se tiene que:

$$\mathbf{A}\underline{Z} = \underline{V}_1 \sim N_p\left(\mathbf{A}\underline{\mu}_{\underline{Z}} \ , \ \mathbf{A} \ _{\underline{Z}}\mathbf{A}^t\right)$$

donde,

$$\mathbf{A}\underline{\mu}_{\underline{Z}} = [c_1\mathbf{I} \quad | \quad c_2\mathbf{I} \quad | \quad \cdots \quad | \quad c_n\mathbf{I}] \begin{bmatrix} \underline{\mu}_1 \\ \underline{\mu}_2 \\ \vdots \\ \vdots \\ \underline{\mu}_n \end{bmatrix} = \sum_{i=1}^n c_i \underline{\mu}_i$$

y $\quad \mathbf{A}\ {}_{\underline{Z}}\mathbf{A}^t =$

$$= [c_1\mathbf{I} \quad | \quad c_2\mathbf{I} \quad | \quad \cdots \quad | \quad c_n\mathbf{I}] \begin{bmatrix} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \end{bmatrix} \begin{bmatrix} c_1\mathbf{I} \\ c_2\mathbf{I} \\ \vdots \\ \vdots \\ c_n\mathbf{I} \end{bmatrix}$$

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

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

para la parte (b):

Sea

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

entonces

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

y de resultados anteriores se tiene que:

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

donde,

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

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

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

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

es decir que,

$$
\underline{\mathbf{v}} = \begin{bmatrix} \underline{\mathbf{v}}_1 \\ \cdots \\ \underline{\mathbf{v}}_2 \end{bmatrix} \sim N_{2p}\left[\begin{pmatrix} \sum_{i=1}^n c_{i-i} \\ \cdots\cdots \\ \sum_{i=1}^n b_{i-i} \end{pmatrix} \;,\; {}_{\underline{\mathbf{v}}_1,\underline{\mathbf{v}}_2}\right]
$$

para la parte (c): Es evidente por resultado de la propiedad 5b (2.7).

Ejemplo 2.10 (Distribuciones de CL de Vectores Aleatorios)**.** Suponga que $\underline{\mathbf{x}}_1, \underline{\mathbf{x}}_2, \underline{\mathbf{x}}_3, \underline{\mathbf{x}}_4$, son vectores aleatorios 3-variados independientes e idénticamente distribuidos, ie. $\underline{\mathbf{x}}_i = (X_1, X_2, X_3)^t$, con:

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

1. Hallar la media y varianza de:
$$\underline{a}^t\underline{\mathbf{x}}_1 = a_1X_1 + a_2X_2 + a_3X_3.$$

2. Hallar la media y varianza de:
$$c_1\underline{\mathbf{x}}_1 + c_2\underline{\mathbf{x}}_2 + c_3\underline{\mathbf{x}}_3 + c_4\underline{\mathbf{x}}_4 = \frac{1}{2}\underline{\mathbf{x}}_1 + \frac{1}{2}\underline{\mathbf{x}}_2 + \frac{1}{2}\underline{\mathbf{x}}_3 + \frac{1}{2}\underline{\mathbf{x}}_4$$

3. Hallar la media y varianza de:
$$\begin{bmatrix} c_1\underline{\mathbf{x}}_1 + c_2\underline{\mathbf{x}}_2 + c_3\underline{\mathbf{x}}_3 + c_4\underline{\mathbf{x}}_4 \\ \underline{\mathbf{x}}_1 + \underline{\mathbf{x}}_2 + \underline{\mathbf{x}}_3 - 3\underline{\mathbf{x}}_4. \end{bmatrix}$$

Solución.

1. La distribución de:
$$\underline{a}^t\underline{\mathbf{x}}_1 = a_1X_1 + a_2X_2 + a_3X_3.$$

   Una combinación lineal de las componentes de un vector aleatorio, ie. una c.l de variables, lo cual es solo una variable aleatoria, por lo tanto se tiene que:
$$
\begin{aligned}
E[\underline{a}^t\underline{\mathbf{x}}_1] &= E[a_1X_1 + a_2X_2 + a_3X_3] \\
&= a_1E[X_1] + a_2E[X_2] + a_3E[X_3] \\
&= 3a_1 - a_2 + a_3,
\end{aligned}
$$

y

$$\begin{aligned}Var[\underline{a}^t\underline{\mathbf{x}}_1] &= Var[a_1X_1 + a_2X_2 + a_3X_3]\\ &= a_1^2Var[X_1] + a_2^2Var[X_2] + a_3^2Var[X_3]\\ &+ 2a_1a_2Cov(X_1, X_2) + 2a_1a_3Cov(X_1, X_3) + 2a_2a_3Cov(X_2, X_3)\\ &= 3a_1^2 + a_2^2 + 2a_3^2 - 2a_1a_2 + 2a_1a_3\\ &= \underline{a}^t\ \underline{a}.\end{aligned}$$

2. La media y varianza de:

$$c_1\underline{\mathbf{x}}_1 + c_2\underline{\mathbf{x}}_2 + c_3\underline{\mathbf{x}}_3 + c_4\underline{\mathbf{x}}_4 = \frac{1}{2}\underline{\mathbf{x}}_1 + \frac{1}{2}\underline{\mathbf{x}}_2 + \frac{1}{2}\underline{\mathbf{x}}_3 + \frac{1}{2}\underline{\mathbf{x}}_4$$

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

$$\begin{aligned}E[c_1\underline{\mathbf{x}}_1 + c_2\underline{\mathbf{x}}_2 + c_3\underline{\mathbf{x}}_3 + c_4\underline{\mathbf{x}}_4] &= c_1E[\underline{\mathbf{x}}_1] + c_2E[\underline{\mathbf{x}}_2] + c_3E[\underline{\mathbf{x}}_3] + c_4E[\underline{\mathbf{x}}_4]\\ &= \sum_{i=1}^{n} c_i\underline{\mu}_i\\ &= c_1\underline{\mu} + c_2\underline{\mu} + c_3\underline{\mu} + c_4\underline{\mu}\\ &= (c_1 + c_2 + c_3 + c_4)\underline{\mu}\\ &= \left(\frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2}\right)\underline{\mu}\\ &= 2\underline{\mu} = 2\begin{bmatrix}3\\-1\\1\end{bmatrix} = \begin{bmatrix}6\\-2\\2\end{bmatrix}\end{aligned}$$

$$\begin{aligned}Var[c_1\underline{\mathbf{x}}_1 + c_2\underline{\mathbf{x}}_2 + c_3\underline{\mathbf{x}}_3 + c_4\underline{\mathbf{x}}_4] &= c_1^2Var[\underline{\mathbf{x}}_1] + c_2^2Var[\underline{\mathbf{x}}_2] + c_3^2Var[\underline{\mathbf{x}}_3] + c_4^2Var[\underline{\mathbf{x}}_4]\\ &= c_1^2\ + c_2^2\ + c_3^2\ + c_4^2\\ &= \left(\sum_{i=1}^{n} c_i^2\right)\\ &= (c_1^2 + c_2^2 + c_3^2 + c_4^2)\\ &= \left(\frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4}\right)\\ &= 1\times\\ &=\\ &= \begin{bmatrix}3 & -1 & 1\\-1 & 1 & 0\\1 & 0 & 2\end{bmatrix}\end{aligned}$$

3. La media y varianza de:

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

Sea
$\underline{\mathbf{v}} = \begin{bmatrix} \underline{\mathbf{v}}_1 \\ \\ \underline{\mathbf{v}}_2 \end{bmatrix}$, luego:

$$Var[\underline{\mathbf{v}}] = \begin{pmatrix} Var(\underline{\mathbf{v}}_1) & | & Cov(\underline{\mathbf{v}}_1\ ,\ \underline{\mathbf{v}}_2) \\ -- & -- & -- \\ Cov(\underline{\mathbf{v}}_2\ ,\ \underline{\mathbf{v}}_1) & | & Var(\underline{\mathbf{v}}_2) \end{pmatrix}$$

$$= \begin{bmatrix} \left(\sum_{i=1}^n c_i^2\right) & \vdots & \left(\sum_{i=1}^n b_i c_i\right) \\ \cdots\cdots & & \cdots\cdots \\ \left(\sum_{i=1}^n b_i c_i\right) & \vdots & \left(\sum_{i=1}^n b_i^2\right) \end{bmatrix}$$

$$= \begin{bmatrix} (\underline{c}^t \underline{c}) & \vdots & \left(\underline{b}^t \underline{c}\right) \\ \cdots\cdots & & \cdots\cdots \\ \left(\underline{b}^t \underline{c}\right) & \vdots & \left(\underline{b}^t \underline{b}\right) \end{bmatrix}_{6\times 6}$$

$$= \begin{bmatrix} 1\times & \vdots & 0\times \\ \cdots\cdots & & \cdots\cdots \\ 0\times & \vdots & 12\times \end{bmatrix} = \begin{bmatrix} & \vdots & \mathbf{0} \\ \cdots\cdots & & \cdots\cdots \\ \mathbf{0} & \vdots & 12 \end{bmatrix}$$

es decir,

$$Var[\underline{\mathbf{v}}] = \begin{bmatrix} & \vdots & \mathbf{0} \\ \cdots\cdots & & \cdots\cdots \\ \mathbf{0} & \vdots & 12 \end{bmatrix} = \begin{bmatrix} 3 & -1 & 1 & & 0 & 0 & 0 \\ -1 & 1 & 0 & \vdots & 0 & 0 & 0 \\ 1 & 0 & 2 & & 0 & 0 & 0 \\ & \cdots\cdots & & & & \cdots\cdots & \\ 0 & 0 & 0 & & 36 & -12 & 12 \\ 0 & 0 & 0 & \vdots & -12 & 12 & 0 \\ 0 & 0 & 0 & & 12 & 0 & 24 \end{bmatrix}$$

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.

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

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

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

$$\underline{\mathbf{v}} = \begin{bmatrix} \underline{\mathbf{v}}_1 \\ \\ \underline{\mathbf{v}}_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\left(X_{11} + X_{21} + X_{31} + X_{41}\right) \\ \\ \frac{1}{2}\left(X_{12} + X_{22} + X_{32} + X_{42}\right) \\ \\ \frac{1}{2}\left(X_{13} + X_{23} + X_{33} + X_{43}\right) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ X_{11} + X_{21} + X_{31} - 3X_{41} \\ \\ X_{12} + X_{22} + X_{32} - 3X_{42} \\ \\ X_{13} + X_{23} + X_{33} - 3X_{43} \end{bmatrix}$$

$$\underline{\mathbf{x}}_i = \begin{bmatrix} 3 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix}$$

$$\underline{\mathbf{v}} = \begin{bmatrix} 3 & -1 & 1 & & 0 & 0 & 0 \\ -1 & 1 & 0 & \vdots & 0 & 0 & 0 \\ 1 & 0 & 2 & & 0 & 0 & 0 \\ & \cdots\cdots & & & & \cdots\cdots & \\ 0 & 0 & 0 & & 36 & -12 & 12 \\ 0 & 0 & 0 & \vdots & -12 & 12 & 0 \\ 0 & 0 & 0 & & 12 & 0 & 24 \end{bmatrix}$$

2.5 Evaluación del Supuesto de Normalidad Multivariada

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 $\underline{\overline{\mathbf{x}}}$ 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 $\underline{\overline{\mathbf{x}}}$ 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 $\underline{\mathbf{x}}_{p\times 1}$, 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 $\underline{\mathbf{x}}_{p\times 1}$, NO parecida a una normal univariada?
3. ¿Al hacer gráficos de dispersión por pares de componentes de $\underline{\mathbf{x}}_{p\times 1}$, 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 $\underline{\mathbf{x}}_{p\times 1}$ y analizando la normalidad bivariada de pares de componentes de dicho vector.

No es frecuente, en la práctica, 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 and Wichern, 2007], un caso de normalidad univariadas que no implica Normalidad-bi-variada.

2.5.1 Evaluación a nivel marginal (ie. Normalidad Univariada)

gds fdg fdg

Bibliography

Richard A Johnson and Dean W Wichern. *Applied multivariate statistical analysis. 6th*. 2007.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2022. URL https://www.R-project.org/.