Repaso de Álgebra Lineal y Presentación de Datos Multivariados

Universidad Nacional de Colombia

Raúl Alberto Pérez

28/2/2022

Técnicas Multivariadas (TM):

Son un conjunto de métodos estadísticos cuya finalidad es analizar simultáneamente conjuntos de datos que involucran la medición de varias variables para cada **caso o individuo**

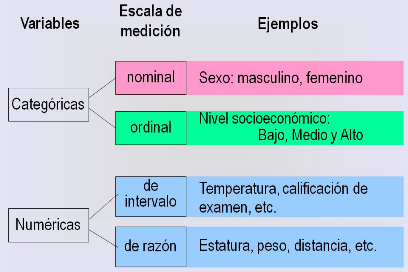
Las TM permiten un mejor entendimiento del fenómeno objeto de estudio, obteniendo información que los métodos univariados y bivariados son incapaces de conseguir.

Objetivos de la TM:

- Proporcionar métodos para estudiar datos multivariados que el análisis estadístico uni y bidimensional es incapaz de conseguir.
- Ayudar al investigador a tomar decisiones óptimas en el contexto en el que se encuentre teniendo en cuenta la información disponible por el conjunto de datos analizado.

Escalas de Medición de las Variables Aleatorias

Los valores asociados a las características de interés relacionadas con el fenómeno de estudio, son observados o medidos en forma diferente, dichas formas se conocen como Niveles o Escalas de Medición. La mayoría de los valores medidos se pueden ubicar en una de la siguiente clasificación:



Clasificación de las TM

Las TM se pueden clasificar en tres grupos:

- Métodos de dependencia
- Métodos de interdependencia
- Métodos estructurales

Métodos de dependencia:

- Suponen que las variables analizadas están divididas en dos grupos: las variables dependientes y las variables independientes.

El objetivo de estos métodos es determinar si el conjunto de variables independientes afecta al conjunto de variables dependientes y de qué forma.

Métodos de interdependencia:

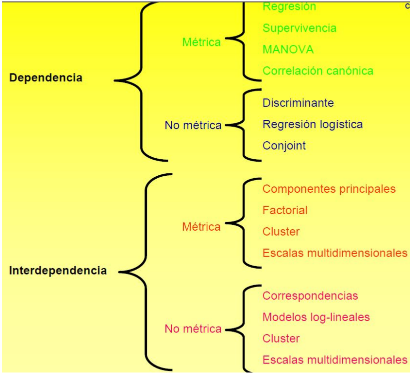
- No distinguen entre variables dependientes e independientes y su objetivo consiste en identificar qué variables están relacionadas, cómo lo están y por qué.

Métodos estructurales:

- Suponen que las variables están divididas en dos grupos: el de las variables dependientes y el de las variables independientes
- Estos métodos tienen aspectos tanto de dependencia como de independencia.

El objetivo de estos métodos es analizar el modelamiento que permite descomponer las relaciones entre las variables, a través de un sistema de ecuaciones lineales, al igual que analizar las pruebas de causalidad involucradas en la variables medidas y en las variables latentes.

Clasificación de las TM:

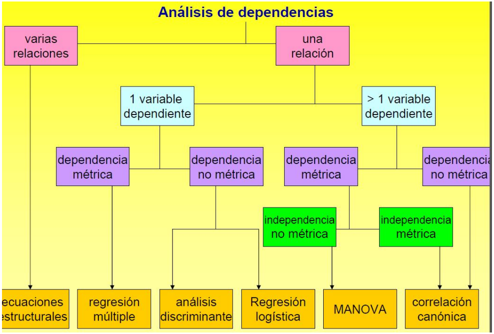


Preguntas:

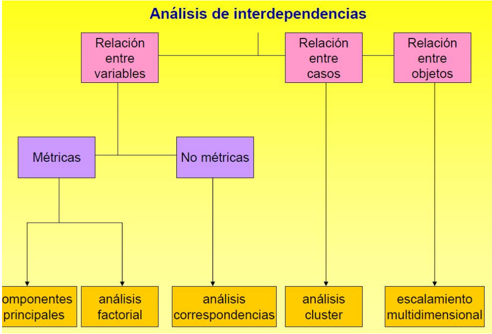
Algunas preguntas a responder antes de iniciar un estudio que involucre la utilización de TM.

- ¿La investigación responde a un problema de dependencia entre variables o de interdependencia de las mismas?
- ¿Cómo están medidas las variables: métricas o no métricas?
- Si es un problema de dependencias, ¿cuántas variables dependientes existen?

Clasificación de las TM de Dependencia:



Clasificación de las TM de Interdependencia:



Repaso de Álgebra Lineal

Algunos conceptos Básicos

1. **Vector:** Arreglo de n-números reales. Se denota por: $\underline{\boldsymbol{x}}$.
2. **Vector de unos:** Vector cuyas entradas son todos el número uno. Se denota por: $\underline{\mathbf{1}}$.
3. **Suma de vectores:** La suma de vectores se realiza componente a componente.
4. **Producto interno entre vectores:** El producto interno entre vectores se realza mediante la multiplicación elemento a elemento, ie.

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

5. **Norma de un Vector:** La norma de un vector es la raíz cuadrada del producto interno del vector por si mismo, ie.

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

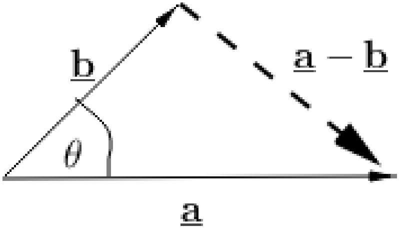
con $\underline{a}^t = (a_1, a_2, \ldots, a_n)$.

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.

6. **Distancia entre dos vectores:** La distancia entre dos vectores es la norma del vector diferencia, ie.

$$d(\underline{a}\ ,\ \underline{b}) = \|\underline{a} - \underline{b}\|$$



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

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

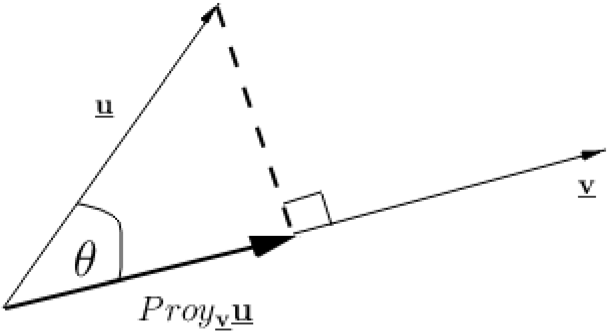
8. **Vectores ortogonales:** Dos vectores $\underline{a}$ y $\underline{b}$ son ortogonales si:

$$\langle \underline{a}\ ,\ \underline{b} \rangle = \underline{a}^t \underline{b} = 0$$

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

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

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



10. **Matriz:** Una matriz es un arreglo rectangular de números en filas y columnas. Se denotan por $\boldsymbol{A}_{n\times p}$
11. **Matriz Cuadrada:** La matriz $\boldsymbol{A}_{n\times p}$ es cuadrada si $n = p$, ie. si el número de filas es igual al número de columnas, se denota por: $\boldsymbol{A}_n$
12. **Matriz Diagonal:** Es una matriz cuadrada con ceros fuera de la diagonal, se denota por $\boldsymbol{D}_n$
13. **Matriz Identidad:** Es una matriz diagonal con unos en la diagonal, se denota por $\boldsymbol{I}_n$
14. **Matriz Cuadrada de unos:** Es una matriz cuadrada llena de unos, se denota por: $\boldsymbol{J}_n$
15. **Matriz Triangular Inferior y Triangular Superior:** Son matrices con parte superior llena de ceros y parte inferior llena de ceros, respectivamente.
16. **Suma de Matrices:** La suma de matrices de igual dimensión se realiza componente a componente.

17. **Multiplicación de Matrices:** La multiplicación de matrices de dimensiones
apropiadas se realiza siguiendo el procedimiento de filas por columnas.
18. **Matriz Transpuesta:** La transpuesta de una matriz $\boldsymbol{A}_{n\times p}$ es la matriz que se obtiene invirtiendo las filas por columnas, la cual se denota por: $\boldsymbol{A}^t_{p\times n}$. Se cumple que: $(\boldsymbol{AB})^t = \boldsymbol{B}^t\boldsymbol{A}^t$.
19. **Matriz Simétrica:** Una matriz cuadrada $\boldsymbol{A}_n$ es simétrica si su transpuesta es igual a la matriz original, ie. si $\boldsymbol{A}^t = \boldsymbol{A}$.
20. **Determinante de una Matriz:** Sea $\boldsymbol{A}_n$ una matriz cuadrada. El determinante de $\boldsymbol{A}$, denotado por $|\boldsymbol{A}|$, se obtiene como:

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

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

Algunas propiedades del determinante de una matriz:

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

a) $|\boldsymbol{A}| = |\boldsymbol{A}^t|$

b) $|\boldsymbol{AB}| = |\boldsymbol{A}||\boldsymbol{B}|$

c) $|c\boldsymbol{A}| = c^n|\boldsymbol{A}|$

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

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

f) Si $\boldsymbol{A}$ es una matriz cuadrada de orden 2-invertible, entonces:

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

21. **Matriz Inversa:** Sea $\boldsymbol{A}_n$ una matriz cuadrada. Si existe una matriz $\boldsymbol{B}_n$ tal que:

$$\boldsymbol{A}_n\boldsymbol{B}_n = \boldsymbol{B}_n\boldsymbol{A}_n = \boldsymbol{I}_n$$

entonces se dice que $\boldsymbol{B}$ es la inversa de $\boldsymbol{A}$ y se denota por: $\boldsymbol{B} = \boldsymbol{A}^{-1}$, ie.

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n$$

Propiedades de la Inversa:

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

1. $(\boldsymbol{A}^t)^{-1} = (\boldsymbol{A}^{-1})^t$
2. $(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$
3. $(\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}$

22. **Matriz Ortogonal:** Una matriz cuadrada $\boldsymbol{A}_n$, se dice que es ortogonal si
su inversa es igual a su transpuesta, ie. se cumple que:

$$\boldsymbol{A}\boldsymbol{A}^t = \boldsymbol{A}^t\boldsymbol{A} = \boldsymbol{I}_n$$

Propiedades de una matriz Ortogonal:

1. $|\boldsymbol{A}| = \pm 1$
2. El producto de un número finito de matrices ortogonales es ortogonal.
3. La Inversa (y por tanto la transpuesta) de una matriz ortogonal es ortogonal.
4. Dada una matriz $\boldsymbol{A}$ y una matriz *ortogonal* $\boldsymbol{P}$, entonces: $|\boldsymbol{A}| = |\boldsymbol{P}^t\boldsymbol{A}\boldsymbol{P}|$

23. **Vectores Linealmente-Independientes (LI):** Un conjunto de k-vectores, $\underline{\boldsymbol{x}}_1, \underline{\boldsymbol{x}}_2, \ldots, \underline{\boldsymbol{x}}_k$,

se dice que es LI si la ecuación:

$$c_1\underline{\boldsymbol{x}}_1 + c_2\underline{\boldsymbol{x}}_2 + \ldots + c_k\underline{\boldsymbol{x}}_k = 0$$

tiene como única solución la dada por:
$c_1 = c_2 = \ldots = c_k = 0 \in \mathbb{R}$.

24. **Valores y Vectores propios de una Matriz:**

Sea $\boldsymbol{A}_n$ una matriz cuadrada. Se dice que λ es un valor propio de $\boldsymbol{A}$, asociado al vector propio $\underline{\boldsymbol{x}}$, si se cumple que: $\boldsymbol{A}\underline{\boldsymbol{x}} = \lambda\underline{\boldsymbol{x}}$.

Para hallar los valores propios de una matriz $\boldsymbol{A}$ se resuelve la siguiente ecuación (llamada ecuación) característica:

$$|\boldsymbol{A} - \lambda\boldsymbol{I}| = 0$$

Al dividir el vector propio $\underline{\boldsymbol{x}}$ por su norma $\|\underline{\boldsymbol{x}}\|$, se obtiene un vector unitario, el cual se denotará por:

$$\underline{\boldsymbol{e}} = \frac{\underline{\boldsymbol{x}}}{\|\underline{\boldsymbol{x}}\|} = \frac{\underline{\boldsymbol{x}}}{\sqrt{\langle \underline{\boldsymbol{x}}\ ,\ \underline{\boldsymbol{x}} \rangle}} = \frac{\underline{\boldsymbol{x}}}{\sqrt{\underline{\boldsymbol{x}}^t\underline{\boldsymbol{x}}}}$$

Alguna propiedades de los valores propios de una matriz:

Sean $\boldsymbol{A}_n$ y $\boldsymbol{B}_n$ dos matrices cuadradas y $c \in \mathbb{R}$, entonces:

1. La matriz $\boldsymbol{A}$ tiene al menos *un valor-propio igual a cero* sii. $\boldsymbol{A}$ es singular (no tiene inversa), ie. sii. el $|\boldsymbol{A}| = 0$.
2. Si $\boldsymbol{A}$ es una matriz simétrica con valores propios reales, entonces los vectores propios asociados a valores propios diferentes son ortogonales.
3. **Teorema de Descomposición Espectral**

Cualquier matriz simétrica $\boldsymbol{A}$, se puede escribir como:

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^t = \boldsymbol{P}^t\boldsymbol{\Lambda}\boldsymbol{P}$$

en donde: $\boldsymbol{\Lambda}$-es una matriz diagonal formada por los valores propios de $\boldsymbol{A}$, y $\boldsymbol{P}$-es una matriz ortogonal cuyas columnas son los vectores propios unitarios asociados a los elementos de la diagonal de $\boldsymbol{\Lambda}$, ie.

$$\boldsymbol{P}\boldsymbol{P}^t = \boldsymbol{P}^t\boldsymbol{P} = \boldsymbol{I}_n$$

Ejemplo: Sea, $\boldsymbol{A} = \begin{bmatrix} 2.2 & 0.4 \\ 0.4 & 2.8 \end{bmatrix}$, luego

$$|\boldsymbol{A} - \lambda \boldsymbol{I}| = \begin{vmatrix} 2.2 - \lambda & 0.4 \\ 0.4 & 2.8 - \lambda \end{vmatrix}$$
$$= \lambda^2 - 5\lambda + 6.16 - 0.16 = (\lambda - 3)(\lambda - 2)$$

luego los valores propios de $\boldsymbol{A}$ son $\lambda_1 = 2$ y $\lambda_2 = 2$.

Ahora los vectores propios se hallan resolviendo el sistema de ecuaciones:

$$\boldsymbol{A}\underline{\boldsymbol{x}}_1 = \lambda_1 \underline{\boldsymbol{x}}_1 \quad \text{y} \quad \boldsymbol{A}\underline{\boldsymbol{x}}_2 = \lambda_2 \underline{\boldsymbol{x}}_2$$

cuya solución es:

$$\underline{\boldsymbol{x}}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{y} \quad \underline{\boldsymbol{x}}_2 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

y dichos vectores propios normalizados son:

$$\underline{\boldsymbol{e}}_1 = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} \quad \text{y} \quad \underline{\boldsymbol{e}}_2 = \begin{bmatrix} 2/\sqrt{5} \\ -1/\sqrt{5} \end{bmatrix}$$

Luego se tiene que: $\boldsymbol{\Lambda} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$ y $\boldsymbol{P} = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{bmatrix}$

Claramente, $\boldsymbol{P}\boldsymbol{P}^t = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Luego del Teorema de la Descomposición espectral se tiene que:

$$\begin{aligned} \boldsymbol{A} &= \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^t \\ &= \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{bmatrix} \\ &= \begin{bmatrix} 3/\sqrt{5} & 4/\sqrt{5} \\ 6/\sqrt{5} & -2/\sqrt{5} \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{bmatrix} \\ &= \begin{bmatrix} 11/5 & 2/5 \\ 2/5 & 14/5 \end{bmatrix} = \begin{bmatrix} 2.2 & 0.4 \\ 0.4 & 2.8 \end{bmatrix} = \boldsymbol{A} \end{aligned}$$

Mas propiedades:

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

$$r(\boldsymbol{A}) = 2 \ , \ (\lambda_1, \lambda_2 \neq 0).$$

5. Si $\lambda_1, \lambda_2, \ldots, \lambda_n$, son los valores propios de $\boldsymbol{A}$, entonces:

$$tr(\boldsymbol{A}) = \sum_{i=1}^{n} \lambda_i \ , \quad \text{y} \quad |\boldsymbol{A}| = \prod_{i=1}^{n} \lambda_i \ , \quad \text{del ejemplo,}$$

$$tr(\boldsymbol{A}) = 2.2 + 2.8 = 5 = 3 + 2 = \lambda_1 + \lambda_2 \ , \quad \text{y}$$

$$|\boldsymbol{A}| = (2.2)(2.8) - (0.4)(0.4) = 6.16 - 0.16 = 6 = 3 \times 2 = \lambda_1 \times \lambda_2$$

6. Si λ es una valor propio de $\boldsymbol{A}$, enton λ^k es un valor propio de $\boldsymbol{A}^k$
7. Las matrices $\boldsymbol{A}$ y $\boldsymbol{A}^t$ tienen el mismo conjunto de valores propios, pero un vector propio de $\boldsymbol{A}$ no-necesariamente es un vector propio de $\boldsymbol{A}^t$.
8. Si $\boldsymbol{A}$ es idempotente, entonces sus valores propios son cero o uno.

25. **Traza de una Matriz:** La traza de una matriz cuadrada $\boldsymbol{A}$, se define como la suma de los elementos de su diagonal, ie.

$$tr(\boldsymbol{A}) = \sum_{i=1}^{n} a_{ii}$$

Alguna propiedades de la traza de una matriz:

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

1. $tr(c\boldsymbol{A}) = c\ tr(\boldsymbol{A})$
2. $tr(\boldsymbol{A} \pm \boldsymbol{B}) = tr(\boldsymbol{A}) \pm tr(\boldsymbol{B})$
3. $tr(\boldsymbol{AB}) = tr(\boldsymbol{BA})$
4. $tr(\boldsymbol{A}^t) = tr(\boldsymbol{A})$
5. $tr(\boldsymbol{AA}^t) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2$
6. $tr(\boldsymbol{ABC}) = tr(\boldsymbol{CAB}) = tr(\boldsymbol{BCA})$
7. Si $\boldsymbol{B}^{-1}$-existe, entonces: $tr(\boldsymbol{B}^{-1}\boldsymbol{AB}) = tr(\boldsymbol{A})$

26. Descomposición Espectral de una Matriz:

Sea $\boldsymbol{A}$ una matriz simétrica de orden n, entonces:

$$\boldsymbol{A} = \lambda_1 \underline{\boldsymbol{e}}_1 \underline{\boldsymbol{e}}_1^t + \lambda_2 \underline{\boldsymbol{e}}_2 \underline{\boldsymbol{e}}_2^t + \ldots + \lambda_n \underline{\boldsymbol{e}}_n \underline{\boldsymbol{e}}_n^t = \sum_{i=1}^{n} \lambda_i \underline{\boldsymbol{e}}_i \underline{\boldsymbol{e}}_i^t$$

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

A la descomposición anterior de $\boldsymbol{A}$-se le llama descomposición espectral de $\boldsymbol{A}$.

Del ejemplo anterior:

$$\begin{aligned}
\boldsymbol{A} &= \sum_{i=1}^{2} \lambda_i \underline{\boldsymbol{e}}_i \underline{\boldsymbol{e}}_i^t = \lambda_1 \underline{\boldsymbol{e}}_1 \underline{\boldsymbol{e}}_1^t + \lambda_2 \underline{\boldsymbol{e}}_2 \underline{\boldsymbol{e}}_2^t \\
&= 3 \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} + 2 \begin{bmatrix} 2/\sqrt{5} \\ -1/\sqrt{5} \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \end{bmatrix} \\
&= 3 \begin{bmatrix} 1/5 & 2/5 \\ 2/5 & 4/5 \end{bmatrix} + 2 \begin{bmatrix} 4/5 & -2/5 \\ -2/5 & 1/5 \end{bmatrix} = \begin{bmatrix} 0.6 & 1.2 \\ 1.2 & 2.4 \end{bmatrix} + \begin{bmatrix} 1.6 & -0.8 \\ -0.8 & 0.4 \end{bmatrix} =
\end{aligned}$$

27. **Matriz Definida y Semi-Definida Positiva:**

Sea $\boldsymbol{A}_n$ una matriz simétrica. Se dice que $\boldsymbol{A}$ es semi-definida positiva si para todo vector $\underline{\boldsymbol{x}} \in \boldsymbol{R}^n$ distinto del vector nulo se cumple que:

$$\underline{\boldsymbol{x}}^t \boldsymbol{A} \underline{\boldsymbol{x}} \geq 0$$

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

Propiedades de una Matriz Definida Positiva (DP):

1. Si $\boldsymbol{A}$ 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.
2. Sea $\boldsymbol{A}$ una matriz DP. Como $\boldsymbol{A}$-es invertible y se tiene que por la descomposición espectral:

$$\boldsymbol{A} = \boldsymbol{P\Lambda P}^t = \sum_{i=1}^{n} \lambda_i \underline{\boldsymbol{e}}_i \underline{\boldsymbol{e}}_i^t, \quad \text{entonces}$$

$$\boldsymbol{A}^{-1} = \boldsymbol{P\Lambda}^{-1} \boldsymbol{P}^t = \sum_{i=1}^{n} \frac{1}{\lambda_i} \underline{\boldsymbol{e}}_i \underline{\boldsymbol{e}}_i^t$$

La raíz-cuadrada de $\boldsymbol{A}$ es:

$$\boldsymbol{A}^{1/2} = \boldsymbol{P}\boldsymbol{\Lambda}^{1/2}\boldsymbol{P}^t = \sum_{i=1}^{n} \sqrt{\lambda_i}\,\underline{\boldsymbol{e}}_i\underline{\boldsymbol{e}}_i^t$$

Del Ejemplo:

$$\begin{aligned}
\boldsymbol{A}^{-1} &= \boldsymbol{P}\boldsymbol{\Lambda}^{-1}\boldsymbol{P}^t \\
&= \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{bmatrix} \begin{bmatrix} 1/3 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{bmatrix} \\
&= \begin{bmatrix} 1/3\sqrt{5} & 1/\sqrt{5} \\ 2/3\sqrt{5} & -1/2\sqrt{5} \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
&= \begin{bmatrix} \frac{1}{15} + \frac{2}{5} & \frac{2}{15} - \frac{1}{5} \\ \frac{2}{15} - \frac{1}{5} & \frac{4}{15} + \frac{1}{10} \end{bmatrix} \\
&= \begin{bmatrix} \frac{35}{75} & -\frac{5}{75} \\ -\frac{5}{75} & \frac{55}{150} \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 2.8 & -0.4 \\ -0.4 & 2.2 \end{bmatrix} = \boldsymbol{A}^{-1}
\end{aligned}$$

Del Ejemplo:

$$
\begin{aligned}
\boldsymbol{A}^{1/2} &= \boldsymbol{P}\boldsymbol{\Lambda}^{1/2}\boldsymbol{P}^t \\
&= \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{bmatrix} \\
&= \begin{bmatrix} \sqrt{3}/\sqrt{5} & 2\sqrt{2}/\sqrt{5} \\ 2\sqrt{3}/\sqrt{5} & -\sqrt{2}/\sqrt{5} \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\sqrt{3}}{5} + \frac{4\sqrt{2}}{5} & 2\frac{\sqrt{3}}{5} - 2\frac{\sqrt{2}}{5} \\ 2\frac{\sqrt{3}}{5} - 2\frac{\sqrt{2}}{5} & 4\frac{\sqrt{3}}{5} + \frac{\sqrt{2}}{5} \end{bmatrix}
\end{aligned}
$$

$$
= \begin{bmatrix} \frac{(\sqrt{3}+4\sqrt{2})}{5} & \frac{(2\sqrt{3}-2\sqrt{2})}{5} \\ \frac{(2\sqrt{3}-2\sqrt{2})}{5} & \frac{(4\sqrt{3}+\sqrt{2})}{5} \end{bmatrix} = \frac{1}{5} \begin{bmatrix} \sqrt{3}+4\sqrt{2} & 2\sqrt{3}-2\sqrt{2} \\ 2\sqrt{3}-2\sqrt{2} & 4\sqrt{3}+\sqrt{2} \end{bmatrix}
$$

$$
= \boldsymbol{A}^{1/2}
$$

Propiedades de la Matriz Raíz-Cuadrada:

La matriz raíz-cuadrada dada por:

$$\boldsymbol{A}^{1/2} = \boldsymbol{P}\boldsymbol{\Lambda}^{1/2}\boldsymbol{P}^t = \sum_{i=1}^{n} \sqrt{\lambda_i}\underline{\boldsymbol{e}}_i\underline{\boldsymbol{e}}_i^t$$

tiene las siguientes propiedades:

1. $(\boldsymbol{A}^{1/2})' = \boldsymbol{A}^{1/2}$, ie. $\boldsymbol{A}^{1/2}$-Es Simétrica.
2. $\boldsymbol{A}^{1/2}\boldsymbol{A}^{1/2} = \boldsymbol{A}$
3. $(A^{1/2})^{-1} = \boldsymbol{P}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{P}^t = \sum_{i=1}^{n} \frac{1}{\sqrt{\lambda_i}}\underline{\boldsymbol{e}}_i\underline{\boldsymbol{e}}_i^t$
4. $\boldsymbol{A}^{1/2}\boldsymbol{A}^{-1/2} = \boldsymbol{A}^{-1/2}\boldsymbol{A}^{1/2} = \boldsymbol{I}$, y $\boldsymbol{A}^{-1/2}\boldsymbol{A}^{-1/2} = \boldsymbol{A}^{-1}$

$\boldsymbol{\Lambda}^{-1/2}$-Matriz diagonal con $1/\sqrt{\lambda_i}$ en la diagonal.

$\boldsymbol{\Lambda}^{1/2}$-Matriz diagonal con $\sqrt{\lambda_i}$ en la diagonal.

$\boldsymbol{\Lambda}^{-1}$-Matriz diagonal con $1/\boldsymbol{\lambda_i}$ en la diagonal.

Formas Cuadráticas

Definición: Sea $\boldsymbol{A}_p$ una matriz simétrica y $\underline{\boldsymbol{x}}$ un vector $p \times 1$, entonces a las función:

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

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

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

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

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

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

y si

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

Algunas Propiedades de Formas-Cuadráticas:

1. Si $\boldsymbol{A} > 0$ entonces, todos sus valores propios son positivos.
2. Si $\boldsymbol{A} \geq 0$ entonces, $\lambda_i \geq 0$ para $i = 1, 2, \ldots, p$ y $\lambda_i = 0$ para algún i.
3. Si $\boldsymbol{A} > 0$ entonces, A-es no-singular y en consecuencia $|\boldsymbol{A}| > 0$.
4. Si $\boldsymbol{A} > 0$ entonces, $\boldsymbol{A}^{-1} > 0$.
5. Si $\boldsymbol{A} > 0$ y $\boldsymbol{C}_p$-es una matriz no-singular entonces, $\boldsymbol{C}^t\boldsymbol{A}\boldsymbol{C} > 0$.
6. Sea $\boldsymbol{A}_p$ una matriz simétrica y sea $\underline{\boldsymbol{x}}$-un vector, entonces

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

Diferenciación con Vectores y Matrices

Función Real de Variable Vectorial:

Sea f-una función que asigna a un vector $\underline{x} \in \mathbb{R}^p$ un número real, ie.

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

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

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

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

Ejemplo: Sea

$$\begin{array}{rcccl} f & : & \mathbb{R}^3 & \longrightarrow & \mathbb{R} \\ \underline{x} = (x_1, x_2, x_3) & & & \longrightarrow & y = f(x_1, x_2, x_3) = 2x_1^2 + 3x_2 + x_3 - x_1 x_2 \end{array}$$

luego,

$$\frac{\partial y}{\partial \underline{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \frac{\partial y}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 4x_1 - x_2 \\ 3 - x_1 \\ 1 \end{bmatrix}$$

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:

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

Ejemplo: Sea

$$f(x_1, x_2, x_3) = 2x_1 + 4x_2 - 3x_3 = (2, 4, -3) \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \underline{a}^t \underline{x}$$

$$\frac{\partial f(\underline{x})}{\partial \underline{x}} = \frac{\partial(\underline{a}^t \underline{x})}{\partial \underline{x}} = \underline{a}^t = (2, 4, -3)$$

2. Si $f(\underline{x}) = \underline{x}^t \boldsymbol{A}$ (vector fila), con $\boldsymbol{A}_p$- entonces:

$$\frac{\partial f(\underline{x})}{\partial \underline{x}} = \frac{\partial (\underline{x}^t \boldsymbol{A})}{\partial \underline{x}} = \boldsymbol{A}, \quad \boldsymbol{A}\text{-como columnas.}$$

Si $f(\underline{x}) = \boldsymbol{A}^t \underline{x}$ (vector columna), con $\boldsymbol{A}_p$- entonces:

$$\frac{\partial f(\underline{x})}{\partial \underline{x}} = \frac{\partial (\boldsymbol{A}^t \underline{x})}{\partial \underline{x}} = \boldsymbol{A}, \quad \boldsymbol{A}\text{-como filas.}$$

Ejemplo:

1. Sea

$$f(x_1, x_2) = \begin{bmatrix} 2x_1 + 3x_2 & x_1 + 4x_2 \end{bmatrix} = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 2 & 3 \\ 1 & 4 \end{pmatrix} = \underline{x}^t \boldsymbol{A}$$

$$\frac{\partial f(\underline{x})}{\partial \underline{x}} = \frac{\partial (\underline{x}^t \boldsymbol{A})}{\partial \underline{x}} = \boldsymbol{A} = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$$

2. Sea

$$f(x_1, x_2) = \begin{bmatrix} 2x_1 + 3x_2 \\ x_1 + 4x_2 \end{bmatrix} = \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix}^t \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 & 3 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \boldsymbol{A}^t \underline{x}$$

$$\frac{\partial f(\underline{x})}{\partial \underline{x}} = \frac{\partial(\boldsymbol{A}^t \underline{x})}{\partial \underline{x}} = \boldsymbol{A} = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$$

3. Derivada de una forma-cuadrática:

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

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

$$\frac{\partial f(\underline{x})}{\partial \underline{x}} = \frac{\partial(\underline{x}^t \boldsymbol{A} \underline{x})}{\partial \underline{x}} = 2\boldsymbol{A} \underline{x}.$$

Ejemplo: Sea

$$f(x_1, x_2) = (x_1, x_2)\begin{bmatrix} 2 & 3 \\ 3 & -2 \end{bmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 + 3x_2 & 3x_1 - 2x_2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= 2x_1^2 + 3x_1x_2 + 3x_1x_2 - 2x_2^2 = 2x_1^2 + 6x_1x_2 - 2x_2^2 = \underline{x}^t\boldsymbol{A}\underline{x}$$

Luego,

$$\frac{\partial f(\underline{x})}{\partial \underline{x}} = \frac{\partial(\underline{x}^t\boldsymbol{A}\underline{x})}{\partial \underline{x}} = 2\boldsymbol{A}\underline{x} = 2\begin{bmatrix} 2 & 3 \\ 3 & -2 \end{bmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= \begin{pmatrix} 4x_1 + 6x_2 \\ 6x_1 - 4x_2 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\partial f(\underline{x})}{\partial x_1} \\ \\ \frac{\partial f(\underline{x})}{\partial x_2} \end{pmatrix}$$

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

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

- $$\frac{\partial f(\boldsymbol{X})}{\partial x_{ij}} = \frac{\partial (\boldsymbol{X}^{-1})}{\partial x_{ij}} = -\boldsymbol{X}^{-1}.\frac{\partial \boldsymbol{X}}{\partial x_{ij}}.\boldsymbol{X}^{-1} = -\boldsymbol{X}^{-1}.\Delta_{ij}.\boldsymbol{X}^{-1},$$

si todos los elemntos de $\boldsymbol{X}$-son diferentes (-i-ésima fila por j-ésima columna de $\boldsymbol{A}^{-1}$).

- Si $\boldsymbol{X}$-es simétrica entonces:

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

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

Donde, $\Delta_{ij} = \frac{\partial (\boldsymbol{X})}{\partial x_{ij}}$-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 $f(\boldsymbol{X}) = |\boldsymbol{X}|$:

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

- $\frac{\partial f(\boldsymbol{X})}{\partial x_{ij}} = \frac{\partial |\boldsymbol{X}|}{\partial x_{ij}} = \boldsymbol{X}_{ij}$, en donde: $\boldsymbol{X}_{ij}$-es el cofactor de x_{ij}.
- La matriz de derivadas es:

$$\frac{\partial |\boldsymbol{X}|}{\partial \boldsymbol{X}} = \begin{bmatrix} \ddots & & \\ & (\boldsymbol{X}_{ij}) & \\ & & \ddots \end{bmatrix}$$

- Para matrices simétricas $\boldsymbol{X}$, la matriz de derivadas es:

$$\frac{\partial |\boldsymbol{X}|}{\partial \boldsymbol{X}} = 2\mathrm{adj}(\boldsymbol{X}) - \mathrm{Diag}[\mathrm{adj}(\boldsymbol{X})] = |\boldsymbol{X}|[2\boldsymbol{X}^{-1} - \mathrm{Diag}(\boldsymbol{X}^{-1})].$$

donde, Diag[adj($\boldsymbol{X}$)-es la matriz diagonal de la adjunta de $\boldsymbol{X}$.

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

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

El cofactor $C_{ij} = \boldsymbol{X}_{ij}$ de $\boldsymbol{X}$ se define como:

$$C_{ij} = (-1)^{i+j}|X_{ij}|,$$

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

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

La derivada de la traza de una matriz $\boldsymbol{X}_p$ es:

$$\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}} = \frac{\partial [tr(\boldsymbol{X})]}{\partial \boldsymbol{X}} = I_p,$$

de donde:

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

7. Derivada del logaritmo del determinante de una matriz:

$f(\boldsymbol{X}) = \mathrm{Log}(|\boldsymbol{X}|)$

La derivada del logaritmo del determinante de una matriz $\boldsymbol{X}_p$ es:

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

Teorema: Dada una matríz simétrica definida positiva $\boldsymbol{B}_p$ y un $b > 0$, ent:

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

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

Matriz de Datos

La matriz de datos muestrales se representa por:

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

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

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 $\underline{\mu}$, combinaciones lineales del vector de medias poblacional y la matriz de var-cov poblacional $\boldsymbol{\Sigma}$.

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

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,

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

se tiene que:

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

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

En la matrix de Var-Cov muestrales:

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

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:

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

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.

Representación Gráfica de Observaciones multivariadas

Algunas gráficas importantes son:

1. Histogramas
2. Diagramas de barras
3. Bix-Plot
4. Gráficos de dispersión y matriz de dispersión
5. Gráficos tridimensionales y de contornos
6. Gráficos de estrellas
7. Gráfico de caras de Chernoff
8. Gráficos de normales bivariadas
9. Gráficos de Cluster

EJEMPLO:

Se tienen datos sobre 6 variables para 25 individuos sobre el contenido de mineral en ciertos huesos.

X_1: dradio , X_2: radio, X_3: dhumero, X_4: humero, X_5: dcubito, X_6: cubito

Lectura de los datos

	dradio	radio	dhumero	humero	dcubito	cubito
1	1.103	1.052	2.139	2.238	0.873	0.872
2	0.842	0.859	1.873	1.741	0.59	0.744
3	0.925	0.873	1.887	1.809	0.767	0.713
4	0.857	0.744	1.739	1.547	0.706	0.674
5	.	.	.	.	.	.
6	.	.	.	.	.	.
7	.	.	.	.	.	.
22	0.94	0.85	2.334	2.225	0.757	0.731
23	0.493	0.616	1.037	1.268	0.546	0.615
24	0.835	0.752	1.509	1.422	0.618	0.664
25	0.915	0.936	1.971	1.869	0.869	0.868

Obtención del vector de medias muestral:

	x
dradio	0.8438
radio	0.8183
dhumero	1.7927
humero	1.7348
dcubito	0.7044
cubito	0.6938

Matriz de var-cov muestral:

	dradio	radio	dhumero	humero	dcubito	cubito
dradio	0.0130	0.0104	0.0223	0.0201	0.0091	0.0080
radio	0.0104	0.0114	0.0185	0.0211	0.0085	0.0089
dhumero	0.0223	0.0185	0.0804	0.0668	0.0168	0.0128
humero	0.0201	0.0211	0.0668	0.0695	0.0177	0.0168
dcubito	0.0091	0.0085	0.0168	0.0177	0.0116	0.0081
cubito	0.0080	0.0089	0.0128	0.0168	0.0081	0.0106

Matriz de correlación muestral:

	dradio	radio	dhumero	humero	dcubito	cubito
dradio	1.0000	0.8518	0.6915	0.6683	0.7437	0.6779
radio	0.8518	1.0000	0.6119	0.7491	0.7422	0.8098
dhumero	0.6915	0.6119	1.0000	0.8936	0.5522	0.4402
humero	0.6683	0.7491	0.8936	1.0000	0.6256	0.6188
dcubito	0.7437	0.7422	0.5522	0.6256	1.0000	0.7289
cubito	0.6779	0.8098	0.4402	0.6188	0.7289	1.0000

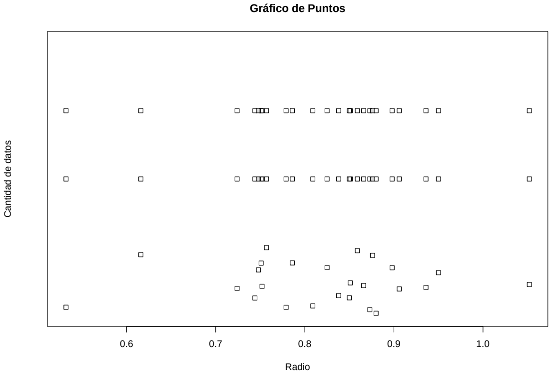
Gráficos

Los gráficos son ayudas importantes en el análisis de los datos. Aunque es imposible graficar simultáneamente los valores de todas las variables en el análisis y estudiar su configuración, los gráficos de las variables individuales y de pares de variables son muy informativos.

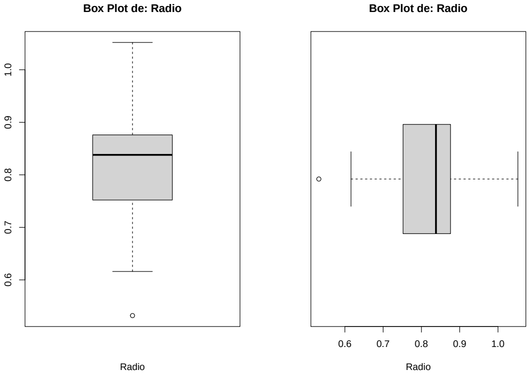
Gráficos para variables individuales:

Sirven para conocer las distribuciones marginales de los datos para cada variable. Entre ellos se encuentran:

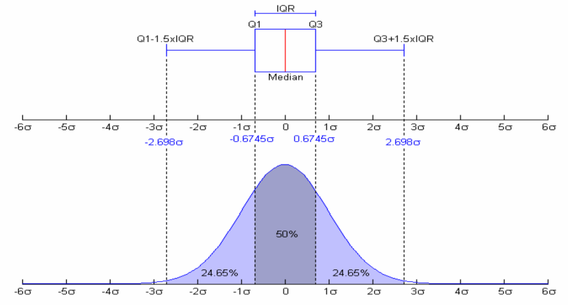
Gráfcio de Puntos: Recomendados para muestras pequeñas.



Gráficos de cajas: Recomendados para muestras moderadas o grandes. Sean Q_1 y Q_3 los cuartiles inferior y superior de la distribución de una variable aleatoria, y sea $IQR = Q_3 - Q_1$ el rango intercuartil. El gráfico de cajas es un gráfico esquemático de la distribución de la variable aleatoria, como se ilustra a continuación.

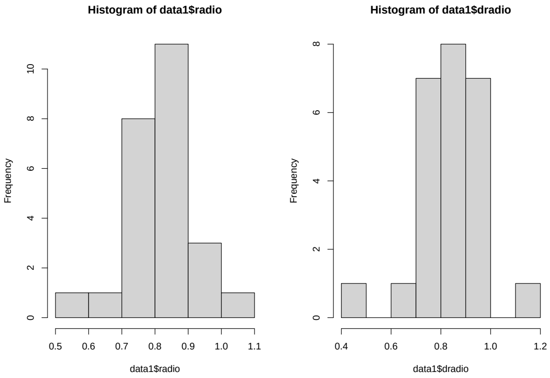


Box-Plot Estándar



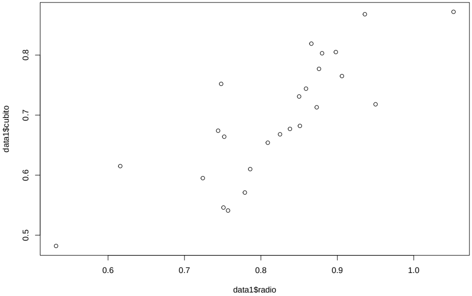
NOTA: Se compara con el caso de que la distribución teórica sea una normal.

Histogramas: Recomendados para muestras moderadas o grandes. Ejemplo, para la variable radio:

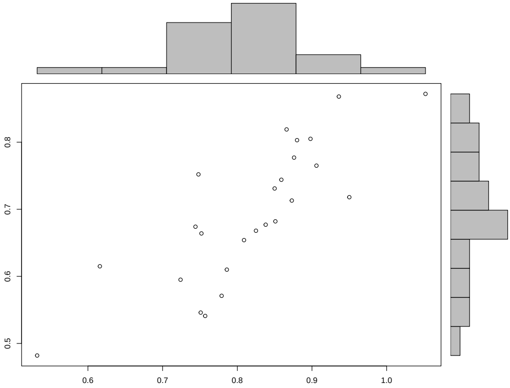


Gráficos para cada par de variables: Son utilizados para estudiar distribución de los datos para 2 variables. Dan indicaciones sobre la orientación de los datos en el plano cartesiano y la asociación que hay entre ellos. También son llamados diagramas de dispersión.

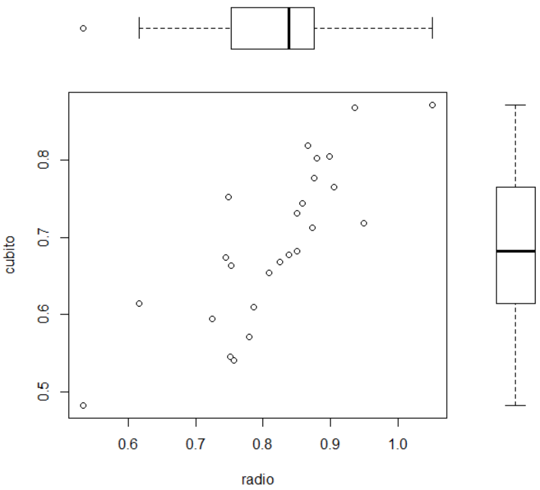
Hay varias clases diagramas de dispersión, por ejemplo:



Diagramas de dispersión simples y con histogramas como marginales:



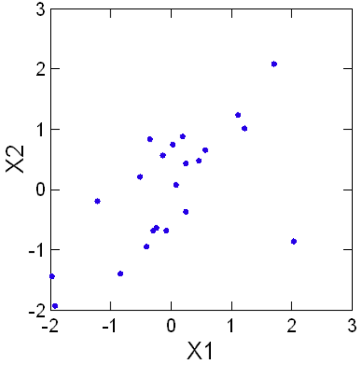
Diagramas de dispersión simples y con box-plot como marginales:



El efecto de observaciones inusuales sobre la correlación muestral:

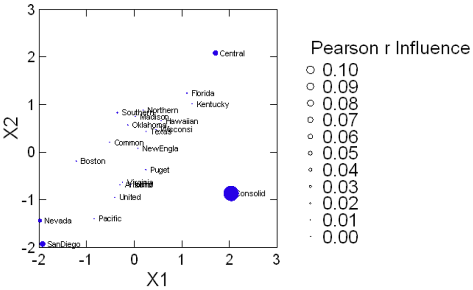
Frecuentemente algunas observaciones de la muestra tienen un efecto considerable en el cálculo de la correlación muestral.

Considere el gráfico de dispersión para dos variables X_1 y X_2.



El coeficiente de correlación muestral es $r_{12} = 0.643$.

Ahora considere el gráfico de dispersión en el cual el tamaño del punto está relacionado con el cambio que tiene el coeficiente de correlación muestral cuando la observación correspondiente a ese punto es eliminada.



Los resultados muestran que al eliminar la observación denominada "consolid", el coeficiente de correlación muestral tiene un cambio mayor de 0.10.

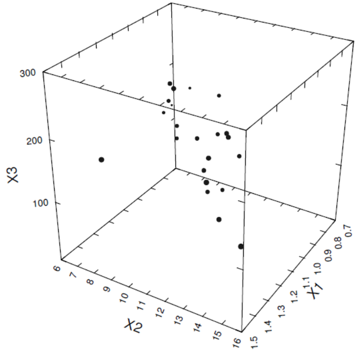
El coeficiente calculado sin esta observación es 0.836. Entonces su eliminación produce un cambio positivo de 0.193 (de 0.643 a 0.836), el cual corresponde a una variación porcentual del 30%.

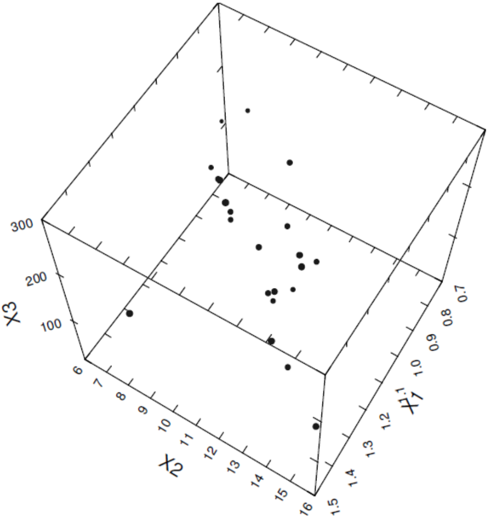
Gráficos para tres variables:

Diagramas de dispersión tridimensionales:

Son utilizados para estudiar los aspectos tridimensionales de los datos. Generalmente estos gráficos permiten rotación.

El siguiente ejemplo presenta el diagrama de dispersión tridimensional para tres variables X_1, X_2 y X_3 con tres rotaciones.



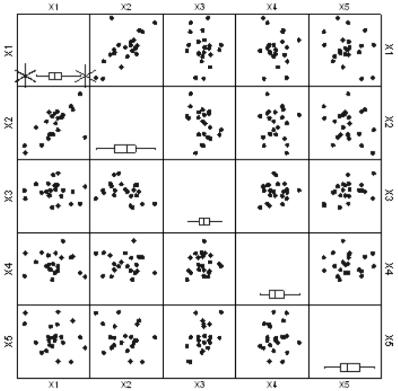


X3
300
200
100
X2
6
7
8
9
10
11
12
13
14
15
16
X1
0.7
0.8
0.9
1.0
1.1
1.2
1.3
1.4
1.5

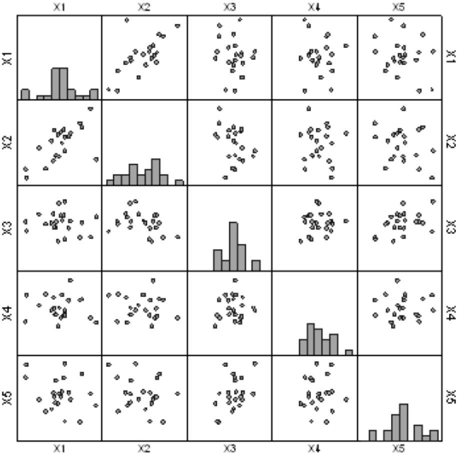
Matrices de dispersión o diagramas múltiples de dispersión

Presentan conjuntamente todos los diagramas de dispersión de los datos para cada par variables. Se pueden construir varias clases de matrices de dispersión, dependiendo del contenido en su diagonal. Por ejemplo:

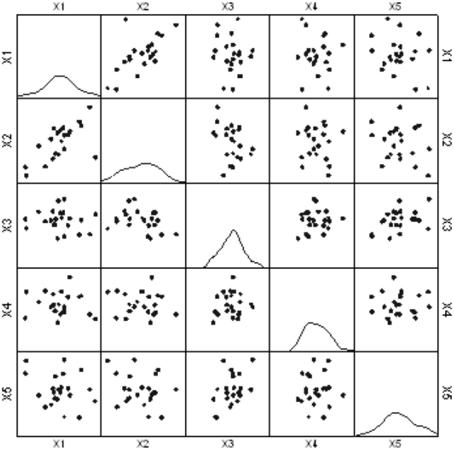
Con diagramas de puntos y con gráficos de cajas en la diagonal:



Con histogramas en la diagonal:



Con histogramas suavizados (curvas Kernel) en la diagonal:



Vectores y Matrices Aleatorias

Vector Aleatorio: Un vector aleatorio es aquel cuyas componentes son variables aleatorias.

Matriz Aleatoria: Una matriz aleatoria es aquella cuyas componentes son variables aleatorias.

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.

Vector de medias poblacional:

Dado un **vector-aleatorio:** $\underline{\boldsymbol{x}} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$ el vector de medias está dado por:

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

Matriz de Var-Cov poblacional:

La matriz de Var-Cov de $\underline{\boldsymbol{x}}$ está dado por:

$$\boldsymbol{\Sigma} := E[(\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}})(\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}})^{'}]$$

$$= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

$$= \begin{bmatrix} Var[X_1] & Cov[X_1, X_2] & \cdots & Cov[X_1, X_p] \\ Cov[X_2, X_1] & Var[X_2] & \cdots & Cov[X_2, X_p] \\ \vdots & \vdots & \ddots & \vdots \\ Cov[X_p, X_1] & Cov[X_p, X_2] & \cdots & Var[X_p] \end{bmatrix}$$

Media Marginal de X_i:

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

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

Varianza Marginal de X_i:

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

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**.

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.

Matriz de Correlación poblacional:

Dado un **vector-aleatorio**

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

la matriz de Correlación de $\underline{\boldsymbol{x}}$ 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 $\boldsymbol{\Sigma}$ y ρ:

Dada la matriz diagonal:

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

se cumple que:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}^{1/2} \rho \boldsymbol{\Lambda}^{1/2}$$

y

$$\rho = (\boldsymbol{\Lambda}^{1/2})^{-1} \boldsymbol{\Sigma} (\boldsymbol{\Lambda}^{1/2})^{-1}$$

Ejemplo:

Hallar la matriz de var-cov del vector aleatorio bi-dimensional $\underline{\boldsymbol{x}}^t = (X_1, X_2)$, 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: $\mu_1 = E[X_1] = 0.1, \quad \mu_2 = E[X_2] = 0.2, \backslash$

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

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

Continuación Ejemplo

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

luego:

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

$$\begin{aligned} \boldsymbol{\Sigma} &= E[(\underline{\boldsymbol{x}} - \underline{\mu})(\underline{\boldsymbol{x}} - \underline{\mu})^t] \\ &= E \begin{bmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 \end{bmatrix} \\ &= \begin{bmatrix} E(X_1 - \mu_1)^2 & E(X_1 - \mu_1)(X_2 - \mu_2) \\ E(X_2 - \mu_2)(X_1 - \mu_1) & E(X_2 - \mu_2)^2 \end{bmatrix} \end{aligned}$$

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

Luego,

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

de donde:

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

y por lo tanto:
$$\begin{aligned} \rho &= (\mathbf{\Lambda}^{1/2})^{-1}\Sigma(\mathbf{\Lambda}^{1/2})^{-1} \\ &= \begin{bmatrix} \frac{1}{\sqrt{0.69}} & \\ & \frac{1}{\sqrt{0.16}} \end{bmatrix} \begin{bmatrix} 0.69 & -0.08 \\ -0.08 & 0.16 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{\sigma_{11}}} & \\ & \frac{1}{\sqrt{\sigma_{22}}} \end{bmatrix} \\ &= \begin{bmatrix} 1 & -0.24 \\ -0.24 & 1 \end{bmatrix} \\ \rho &= \begin{bmatrix} 1 & Corr(X_1, X_2) \\ Corr(X_2, X_1) & 1 \end{bmatrix} \end{aligned}$$

Ejemplo-2:

Suponga que la matriz de Var-Cov del vector-aleatorio tri-dimensional:
$\underline{\boldsymbol{x}} = (X_1, X_2, X_3)^t$ es:

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

Hallar: ρ.

Particionamiento del Vector de Medias-Poblacionales

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

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

Particionamiento de la Matriz de Var-Cov Poblacional $\mathbf{\Sigma}$

$$\mathbf{\Sigma} = \left[\begin{array}{ccccccc} \sigma_{11} & \cdots & \sigma_{1q} & | & \sigma_{1,q+1} & \cdots & \sigma_{1p} \\ \vdots & & \vdots & | & \vdots & & \vdots \\ \sigma_{q1} & \cdots & \sigma_{qq} & | & \sigma_{q,q+1} & \cdots & \sigma_{qp} \\ -- & -- & -- & -- & -- & -- & -- \\ \sigma_{q+1,1} & \cdots & \sigma_{q+1,q} & | & \sigma_{q+1,q+1} & \cdots & \sigma_{q+1,p} \\ \vdots & & \vdots & | & \vdots & & \vdots \\ \sigma_{p,1} & \cdots & \sigma_{p,q} & | & \sigma_{p,q+1} & \cdots & \sigma_{p,p} \end{array}\right]$$

$$= \begin{array}{cc} & \begin{array}{ccc} q & & p-q \end{array} \\ \begin{array}{c} q \\ \\ p-q \end{array} & \left[\begin{array}{ccc} \mathbf{\Sigma}_{11} & | & \mathbf{\Sigma}_{12} \\ -- & -- & -- \\ \mathbf{\Sigma}_{21} & | & \mathbf{\Sigma}_{22} \end{array}\right] \end{array}$$

La anterior matriz particionada se obtiene como sigue:

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

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

Ahora, tomando el valor esperado a ambos lados se tiene:

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

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

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

Similarmente para $\mathbf{\Sigma}_{11}$, $\mathbf{\Sigma}_{22}$ y $\mathbf{\Sigma}_{21} = \mathbf{\Sigma}_{12}^t$

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

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

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

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

$$\text{luego,} \quad \mathbf{\Sigma} = E\left[(\underline{\boldsymbol{x}} - \underline{\mu})(\underline{\boldsymbol{x}} - \underline{\mu})^t\right] = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix}$$

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. $[cX_1] = cE[X_1]$

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

Notar que si, $aX_1 + bX_2 = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \underline{\boldsymbol{c}}^t \underline{\boldsymbol{x}}$, de donde:

$$E[aX_1 + bX_2] = E[\underline{\boldsymbol{c}}^t \underline{\boldsymbol{x}}] = a\mu_1 + b\mu_2 = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \underline{\boldsymbol{c}}^t \underline{\mu}$$

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

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

$$\underline{\boldsymbol{c}}^t \Sigma \underline{\boldsymbol{c}} = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = a^2\sigma_{11} + b^2\sigma_{22} + 2ab\sigma_{12}$$

En resumen:

$$Var\left[aX_1 + bX_2\right] = Var\left(\underline{\boldsymbol{c}}^t \underline{\boldsymbol{x}}\right) = \underline{\boldsymbol{c}}^t \mathbf{\Sigma} \underline{\boldsymbol{c}}$$

En general, dado vector aleatorio $\underline{\boldsymbol{x}} = \left(X_1, X_2, \cdots, X_p\right)^t$, con matriz de Var-Cov dada por $\mathbf{\Sigma}_{p\times p}$, si $\underline{\boldsymbol{c}} = \left(c_1, c_2, \cdots, c_p\right)^t$ es un vector de constantes, entonces:

$$E\left[\underline{\boldsymbol{c}}^t \underline{\boldsymbol{x}}\right] = E\left[c_1X_1 + c_2X_2 + \cdots + c_pX_p\right]$$

$$= \underline{\boldsymbol{c}}^t \underline{\boldsymbol{\mu}},$$

$$Var\left[\underline{\boldsymbol{c}}^t \underline{\boldsymbol{x}}\right] = Var\left[c_1X_1 + c_2X_2 + \cdots + c_pX_p\right] = \underline{\boldsymbol{c}}^t \mathbf{\Sigma} \underline{\boldsymbol{c}}.$$

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

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

$$E\left[\underline{\boldsymbol{z}}\right] = E\left[\boldsymbol{C}\underline{\boldsymbol{x}}\right] = \boldsymbol{C}\underline{\mu}$$

$$Var\left[\underline{\boldsymbol{z}}\right] = Var\left[\boldsymbol{C}\underline{\boldsymbol{x}}\right] = \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^t$$

7. Sea $\boldsymbol{A} = \left[(a_{ij})\right]_{q_1\times p}$ y $\boldsymbol{B} = \left[(b_{ij})\right]_{q_2\times p}$ matrices de constantes y $\underline{\boldsymbol{x}} = (X_1, X_2, \cdots, X_p)^t$ un vector aleatorio con vector de medias $\underline{\mu}$ y matriz de Var-Cov $\boldsymbol{\Sigma}_{p\times p}$. Para los vectores aleatorios: $\underline{\boldsymbol{z}}_1 = \boldsymbol{A}\underline{\boldsymbol{x}}$ y $\underline{\boldsymbol{z}}_2 = \boldsymbol{B}\underline{\boldsymbol{x}}$, se tiene que:

$$Cov\left[\underline{\boldsymbol{z}}_1\ ,\ \underline{\boldsymbol{z}}_2\right] = Cov\left[\boldsymbol{A}\underline{\boldsymbol{x}}\ ,\ \boldsymbol{B}\underline{\boldsymbol{x}}\right] = \underbrace{\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}^t}_{q_1\times q_2}, \quad \text{similarmente}$$

$$Cov\left[\underline{\boldsymbol{z}}_1\ ,\ \underline{\boldsymbol{z}}_2\right] = Cov\left[\boldsymbol{A}\underline{\boldsymbol{x}}\ ,\ \boldsymbol{B}\underline{\boldsymbol{y}}\right] = \underbrace{\boldsymbol{A}\boldsymbol{\Sigma}_{XY}\boldsymbol{B}^t}_{q_1\times q_2}, \text{ con } \underline{\boldsymbol{x}}, \underline{\boldsymbol{y}} - vect - aleats.$$

Ejemplo : Sea $\underline{\boldsymbol{x}} = \begin{pmatrix} X_1 & X_2 \end{pmatrix}^t$-un vector aleatorio con media $\underline{\mu} = \begin{pmatrix} \mu_1 & \mu_2 \end{pmatrix}$ y matriz de Var-Cov dada por: $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$.

Sea el vector aleatorio $\underline{\boldsymbol{z}}_{2\times 1} = \begin{pmatrix} Z_1 & Z_2 \end{pmatrix}^t$ 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 $\underline{\boldsymbol{z}}$.

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

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

luego, usando el resultado anterior se tiene que:

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

$$Var[\underline{\boldsymbol{z}}] = Var[\boldsymbol{C}\underline{\boldsymbol{x}}] = \boldsymbol{C\Sigma C}^t$$

$$= \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} - \sigma_{12} & \sigma_{12} - \sigma_{22} \\ \sigma_{11} + \sigma_{12} & \sigma_{12} + \sigma_{22} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

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

$$= \begin{bmatrix} \sigma_{11} - 2\sigma_{12} + \sigma_{22} & \sigma_{11} - \sigma_{22} \\ \\ \sigma_{11} - \sigma_{22} & \sigma_{11} + 2\sigma_{12} + \sigma_{22} \end{bmatrix} = \begin{bmatrix} Var[Z_1] & Cov[Z_1\ ,\ Z_2] \\ \\ Cov[Z_2\ ,\ Z_1] & Var[Z_2] \end{bmatrix}$$

Particionamiento del Vector de Medias Muestrales

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

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

Particionamiento de la Matriz de Var-Cov Muestrales

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

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

Algunas formas matriciales Eficientes

Para la matriz de datos

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

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. $\underline{\bar{\boldsymbol{x}}} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{bmatrix} = \frac{1}{n}\boldsymbol{X}^T\mathbf{1} = \frac{1}{n}\begin{bmatrix} \sum_{i=1}^n x_{i1} \\ \sum_{i=1}^n x_{i2} \\ \vdots \\ \sum_{i=1}^n x_{ip} \end{bmatrix}_{p\times 1}$
2. $\boldsymbol{S} = \frac{1}{n-1}\boldsymbol{X}^T\left[\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\right]\boldsymbol{X} = \frac{1}{n-1}\boldsymbol{X}^T\boldsymbol{H}\boldsymbol{X}$, con
   $\boldsymbol{H} = \boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$
3. $\boldsymbol{R} = D^{-\frac{1}{2}}\boldsymbol{S}D^{-\frac{1}{2}}$, 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: $\boldsymbol{S} = D^{\frac{1}{2}}\boldsymbol{R}D^{\frac{1}{2}}$

4. Se define la Varianza-Generalizada de $\boldsymbol{X}$ como el determinante de $\boldsymbol{S}$,

$$VG = |\boldsymbol{S}|,$$

la cual representa una medida de variabilidad del vector de variables aleatorias $\underline{\boldsymbol{x}}$

5. Se define la Varianza Total de $\boldsymbol{X}$ como la traza de $\boldsymbol{S}$, ie

$$VT = Tr(\boldsymbol{S})$$

Muestra Aleatoria de Distribuciones p-Variadas

Sea un vector p-variado de variables aleatorias $\underline{\boldsymbol{x}} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$,

con función de distribución multivariada representada por $f(\underline{\boldsymbol{x}})$, con media $\underline{\mu}$ y var-cov $\boldsymbol{\Sigma}$, ie, $E[\underline{\boldsymbol{x}}] = \underline{\mu}$ y $\mathrm{Var}(\underline{\boldsymbol{x}}) = \boldsymbol{\Sigma}$.

Una muestra aleatoria de tamaño n de esta distribución es un conjunto de n-vectores aleatorios p-variados, $\underline{\boldsymbol{x}}_1, \underline{\boldsymbol{x}}_2, \cdots, \underline{\boldsymbol{x}}_n$, independientes e identicamente distribuídos con distribución $f(\underline{\boldsymbol{x}})$, es decir que:

$$h(\underline{\boldsymbol{x}}_1, \underline{\boldsymbol{x}}_2, \cdots, \underline{\boldsymbol{x}}_n) = h_1(\underline{\boldsymbol{x}}_1) h_2(\underline{\boldsymbol{x}}_2) \ldots h_n(\underline{\boldsymbol{x}}_n) = \prod_{i=1}^{n} f(\underline{\boldsymbol{x}}_i)$$

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

Teorema: Propiedades de $\overline{\underline{\boldsymbol{x}}}$

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

1. $E[\overline{\underline{\boldsymbol{x}}}] = \underline{\mu}$
2. Var$[\overline{\underline{\boldsymbol{x}}}]$=$\frac{1}{n}\boldsymbol{\Sigma}$
3. $E[\boldsymbol{S}] = \boldsymbol{\Sigma}$ y $E[\boldsymbol{S}_n] = \left(\frac{n-1}{n}\right)\boldsymbol{\Sigma}$

Dm: Para la parte (i) observe que:

$$\underline{\overline{\boldsymbol{x}}}_{p\times 1} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{bmatrix} = \frac{1}{n}\begin{bmatrix} \sum_{i=1}^n x_{i1} \\ \sum_{i=1}^n x_{i2} \\ \vdots \\ \sum_{i=1}^n x_{ip} \end{bmatrix} = \frac{1}{n}\begin{bmatrix} x_{11}+x_{21}+\cdots+x_{n1} \\ x_{12}+x_{22}+\cdots+x_{n2} \\ \vdots \\ x_{1p}+x_{2p}+\cdots+x_{np} \end{bmatrix}_{p\times 1}$$

$$= 1/n[\underline{\boldsymbol{x}}_1 + \underline{\boldsymbol{x}}_2 + \cdots + \underline{\boldsymbol{x}}_n]$$

Luego de lo anterior se tiene que:

$$E[\underline{\overline{\boldsymbol{x}}}] = \frac{1}{n}E\left[\underline{\boldsymbol{x}}_1 + \underline{\boldsymbol{x}}_2 + \cdots + \underline{\boldsymbol{x}}_n\right]$$

$$= \frac{1}{n}(E[\underline{\boldsymbol{x}}_1] + E[\underline{\boldsymbol{x}}_2] + \cdots + E[\underline{\boldsymbol{x}}_n])$$

$$E[\underline{\overline{\boldsymbol{x}}}] = \frac{1}{n}(\underline{\mu} + \underline{\mu} + \cdots + \underline{\mu}) = \frac{1}{n}n\underline{\mu} = \underline{\mu}$$

Para la parte (ii), oberve que:

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

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

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

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

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

$$= \frac{1}{n^2}\left[\boldsymbol{\Sigma} + \boldsymbol{\Sigma} + \cdots + \boldsymbol{\Sigma} + 0 + 0 + \cdots + 0\right] = \frac{1}{n^2} n\boldsymbol{\Sigma} = \frac{1}{n}\boldsymbol{\Sigma}$$

Para la parte (iii), observe que:

$$\sum_{j=1}^{n}(\underline{\boldsymbol{x}}_j - \overline{\underline{\boldsymbol{x}}})(\underline{\boldsymbol{x}}_j - \overline{\underline{\boldsymbol{x}}})^T = \sum_{j=1}^{n}(\underline{\boldsymbol{x}}_j - \overline{\underline{\boldsymbol{x}}})(\underline{\boldsymbol{x}}_j^T - \overline{\underline{\boldsymbol{x}}}^T)$$

$$= \sum_{j=1}^{n}(\underline{\boldsymbol{x}}_j - \overline{\underline{\boldsymbol{x}}})\underline{\boldsymbol{x}}_j^T - \sum_{j=1}^{n}(\underline{\boldsymbol{x}}_j - \overline{\underline{\boldsymbol{x}}})\overline{\underline{\boldsymbol{x}}}^T$$

$$= \sum_{j=1}^{n}\underline{\boldsymbol{x}}_j\underline{\boldsymbol{x}}_j^T - n\overline{\underline{\boldsymbol{x}}}\ \overline{\underline{\boldsymbol{x}}}^T,$$

lo anterior debido a que se cumple lo siguiente:

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

Tomando valor esperado en lo anterior se tiene que:

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

$$=\sum_{j=1}^{n}E\left[\underline{\boldsymbol{x}}_j\underline{\boldsymbol{x}}_j^T\right]-nE\left[\overline{\underline{\boldsymbol{x}}}\ \overline{\underline{\boldsymbol{x}}}^T\right],$$

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

$$Var(\ \underline{\boldsymbol{x}}\ )=\boldsymbol{\Sigma}=E\left[\underline{\boldsymbol{x}\boldsymbol{x}}^T\right]-E\left[\underline{\boldsymbol{x}}\right]E\left[\underline{\boldsymbol{x}}\right]^T,$$

es decir que,

$$E\left[\underline{\boldsymbol{x}\boldsymbol{x}}^T\right]=\boldsymbol{\Sigma}+E\left[\underline{\boldsymbol{x}}\right]E\left[\underline{\boldsymbol{x}}\right]^T,$$

de donde:

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

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

$$= \sum_{j=1}^{n} \left(\boldsymbol{\Sigma} + \underline{\mu\mu}^T \right) - n\left(\frac{1}{n}\boldsymbol{\Sigma} + \underline{\mu\mu}^T \right)$$

$$= n\boldsymbol{\Sigma} + n\underline{\mu\mu}^T - \boldsymbol{\Sigma} - n\underline{\mu\mu}^T = (n-1)\boldsymbol{\Sigma},$$

es decir que,

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

Pero

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

y por lo tanto se tiene que:

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

$$= \frac{1}{n-1}(n-1)\boldsymbol{\Sigma}$$

$$E[\boldsymbol{S}] = \boldsymbol{\Sigma}$$

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

Distancias

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

Dados dos vectores $\underline{\boldsymbol{x}}, \underline{\boldsymbol{y}} \in \mathbb{R}^p$, $\underline{\boldsymbol{x}} = (x_1, x_2, \ldots, x_p)$ y $\underline{\boldsymbol{y}} = (y_1, y_2, \ldots, y_p)$, la distancia euclídea entre ellos se define como sigue:

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

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.

Distancia Estadística

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

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

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

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

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.

Distancia de Mahalanobis

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

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

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

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

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

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

Ejemplo Práctico:

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.

Datos del ejemplo

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?