Análisis de Datos Multivariantes

1. INTRODUCCIÓN

(Parte 1: Elementos y herramientas básicas)

2016/17

Universidad Carlos III de Madrid

Contenido

Datos Multivariantes

Universidad Carlos III de Madrid

- **Evaluación Profesional** (⋆)

X_0	X_1	X_2	X_3	X_4	X_5	X_6
43	51	30	39	61	92	45
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
82	82	39	59	64	78	39

(⋆) Datos completos en `attitude.txt`.

X_0: **Evaluación global**.

X_1: **Quejas**.
X_2: **Privilegios**.
X_3: **Oportunidades**.
X_4: **Eficacia**.
X_5: **Críticas**.
X_6: **Promoción**. (%)

Datos Multivariantes

- **Notas de Exámenes con Libro Cerrado-Abierto** $^{(\star)}$

	X_1	X_2	X_3	X_4	X_5
1	77	82	67	67	81
2	63	78	80	70	81
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
87	05	26	15	20	20
88	00	40	21	09	14

$^{(\star)}$ Datos completos en `opencl.txt`.

X_1: **Mecánica** (C).
X_2: **Vectores** (C).
X_3: **Álgebra** (A).
X_4: **Análisis** (A).
X_5: **Estadística** (A). ($[0, 100]$)

Datos Multivariantes

- **Calidad del Papel de Impresora** $^{(\star)}$

X_1	X_2	X_3
0.801	121.41	70.42
0.824	127.70	72.47
⋮	⋮	⋮
0.776	110.71	53.67
0.758	113.80	52.42

$^{(\star)}$ Datos completos en printer.txt.

X_1: **Densidad** (en gr./cm^3).
X_2: **Resistencia longitudinal** (en lb).
X_3: **Resistencia transversal** (en lb).

Matriz de Datos

$\mathcal{X} = (x_{ij}) =$

i	X_0	X_1	X_2	X_3	X_4	X_5	X_6
1	43	51	30	39	61	92	45
2	63	64	51	54	63	73	47
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
29	85	85	71	71	77	74	55
30	82	82	39	59	64	78	39

- $x_{1,1} = 51$, $x_{2,4} = 63$, $x_{29,0} = 85$, ...
- $n = 30$ filas (**objetos**, individuos, unidades experimentales, ...).
- $p = 7(1 + 6)$ columnas (**variables**).
- n: **Tamaño muestral**. p: **Dimensión**.

Matriz de Datos

- Se puede escribir

$$\mathcal{X} = \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix} = \begin{pmatrix} \mathbf{c}_1 \cdots \mathbf{c}_p \end{pmatrix} .$$

  - $\mathbf{x}_i^T$: Mediciones $i-$ésimo **objeto** $(1 \times p)$. Por ejemplo:

$$\mathbf{x}_1^T = \begin{pmatrix} 43 & 51 & 30 & 39 & 61 & 92 & 45 \end{pmatrix} .$$

  - $\mathbf{c}_j$: Mediciones $j-$ésima **variable** $(n \times 1)$. Por ejemplo:

$$\mathbf{c}_2 = \begin{pmatrix} 30 & 51 & \cdots & 71 & 39 \end{pmatrix}^T = \begin{pmatrix} 30 \\ \vdots \\ 39 \end{pmatrix} .$$

1. Datos Multivariantes
   - Ejemplo 1: Evaluación Profesional
   - Ejemplo 2: Exámenes con Libro Cerrado-Abierto
   - Ejemplo 3: Calidad del Papel de Impresora
2. Matriz de Datos
   - Elementos y notación
   - Filas y columnas
3. Estadísticos Resumen
   - Vector de medias muestrales
   - Matriz de covarianzas muestrales
   - Matriz de correlaciones muestrales
   - Expresiones matriciales
4. Representaciones Gráficas
   - Matriz de nubes de puntos
   - Separación de clases

Estadísticos Resumen

Universidad Carlos III de Madrid

- Media muestral de cada columna $\mathbf{c}_j$:

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \ , \quad j = 1, \ldots, p \ .$$

Vector de medias muestrales $(p \times 1)$

$$\bar{\mathbf{x}} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{pmatrix} .$$

- Repetir la computación univariante p veces.

Estadísticos Resumen

- Covarianza muestral entre pares de columnas $\mathbf{c}_j$ y $\mathbf{c}_k$:

$$s_{jk} = \frac{1}{n}\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)(x_{ik}-\bar{x}_k)\ , \quad j,k=1,\ldots,p\ .$$

Matriz de covarianzas muestrales ($p \times p$)

$$\mathbf{S} = \left(\begin{array}{c|cccc} & X_1 & X_2 & \ldots & X_p \\ \hline X_1 & s_{11} & s_{12} & \ldots & s_{1p} \\ X_2 & s_{21} & s_{22} & \ldots & s_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_p & s_{p1} & s_{p2} & \ldots & s_{pp} \\ \hline \end{array}\right) = \begin{pmatrix} s_{11} & s_{12} & \ldots & s_{1p} \\ s_{21} & s_{22} & \ldots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \ldots & s_{pp} \end{pmatrix} .$$

- Repetir la computación bivariante $p(p+1)/2$ veces (?).

Estadísticos Resumen

- Correlación entre pares de columnas $\mathbf{c}_j$ y $\mathbf{c}_k$:

$$r_{jk} = \frac{s_{jk}}{\sqrt{s_{jj}\, s_{kk}}} \ , \quad j,k = 1,\ldots,p \ .$$

Matriz de correlaciones muestrales ($p \times p$)

$$\mathbf{R} = \left(\begin{array}{c|cccc} & X_1 & X_2 & \ldots & X_p \\ \hline X_1 & 1 & r_{12} & \ldots & r_{1p} \\ X_2 & r_{21} & 1 & \ldots & r_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_p & r_{p1} & r_{p2} & \ldots & 1 \end{array} \right) = \begin{pmatrix} 1 & r_{12} & \ldots & r_{1p} \\ r_{21} & 1 & \ldots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \ldots & 1 \end{pmatrix} .$$

Estadísticos Resumen

Universidad Carlos III de Madrid

Expresiones matriciales

Dada una **matriz de datos** $\mathcal{X} = (x_{ij})$ de $n \times p$:

- **Vector de medias muestrales:**

$$\bar{\mathbf{x}} = \frac{1}{n}\mathcal{X}^T \mathbf{1}_n = (\bar{x}_j : j = 1, \ldots, p) \; .$$

- **Matriz de covarianzas muestrales:**

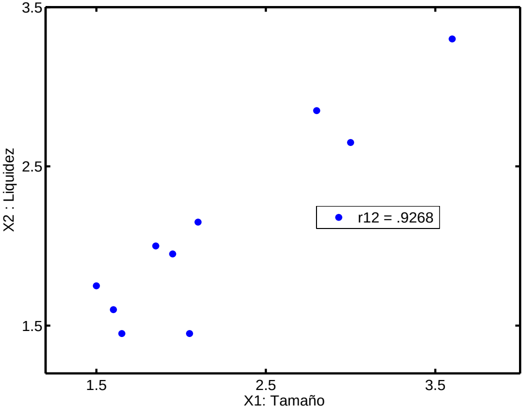
$$\mathbf{S} = \frac{1}{n}\mathcal{X}^T\mathcal{X} - \bar{\mathbf{x}}\,\bar{\mathbf{x}}^T = (s_{jk} : j, k = 1, \ldots, p) \; .$$

- **Matriz de correlaciones muestrales:** Si $\mathbf{D} = \mathrm{diag}(\mathbf{S})$:

$$\mathbf{R} = \mathbf{D}^{-1/2}\mathbf{S}\mathbf{D}^{-1/2} = \mathbf{S} * \mathbf{d}\mathbf{d}^T = \quad [\mathbf{d} = \mathrm{diag}(\mathbf{D}^{-1/2})]$$

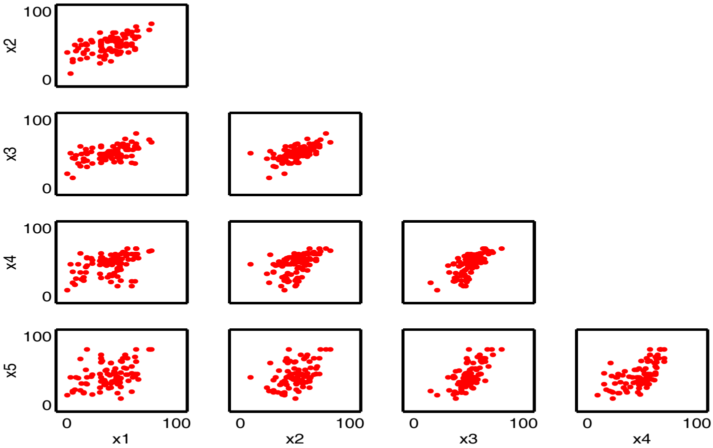
$$= \left(r_{jk} = \frac{s_{jk}}{\sqrt{s_{jj}\, s_{kk}}} : j, k = 1, \ldots, p\right) \; .$$

Representaciones Gráficas



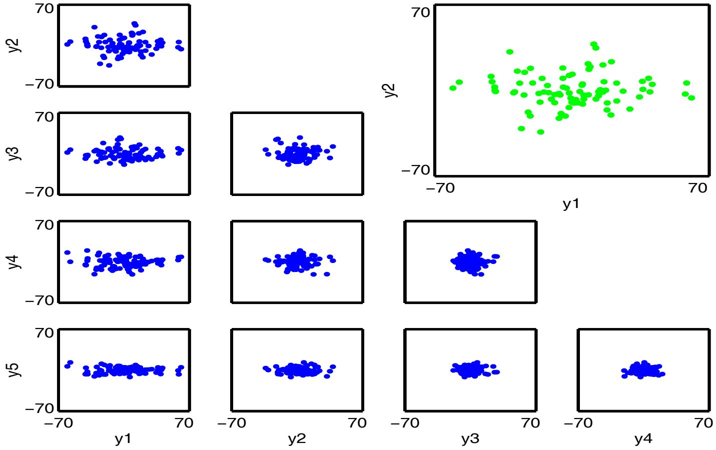
Nube de Puntos (X_1, X_2) (Datos de Empresas)

Representaciones Gráficas



Matriz de Nubes de Puntos (I) (Exámenes con Libro Cerrado-Abierto)

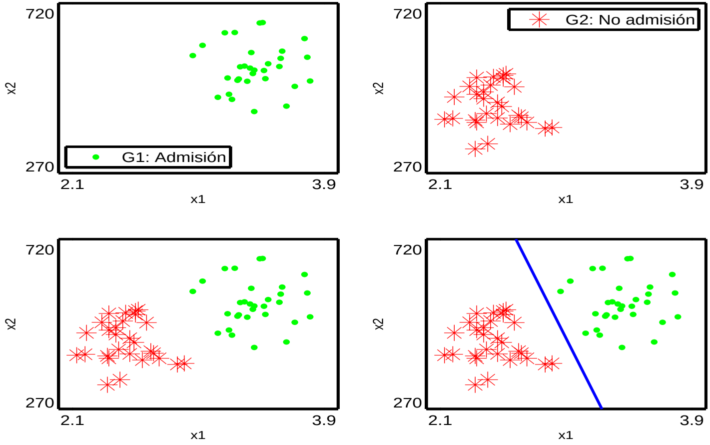
Representaciones Gráficas



Matriz de Nubes de Puntos (II) (Exámenes con Libro Cerrado-Abierto)

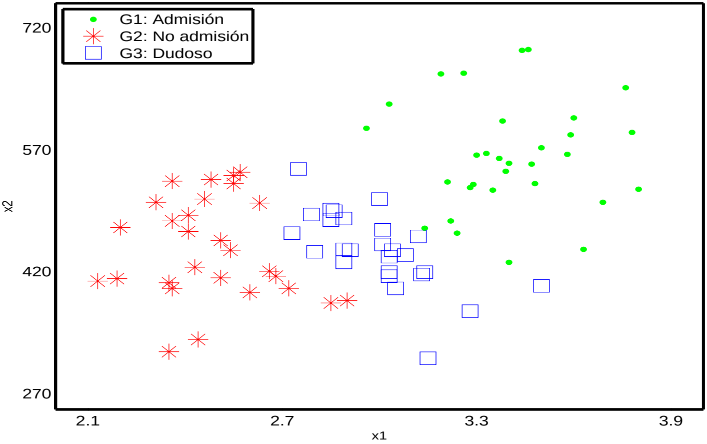
Representaciones Gráficas

Universidad Carlos III de Madrid



Dos grupos (Datos de Admisión)

Representaciones Gráficas



Tres grupos (Datos de Admisión)

Resumen

Universidad Carlos III de Madrid

1. Datos Multivariantes
2. Matriz de Datos
3. Estadísticos Resumen
4. Representaciones Gráficas

- **Referencias**: Johnson, R.A. y Wichern, D.W. (2007) [Cap. 1].

Análisis de Datos Multivariantes

1. INTRODUCCIÓN

(Parte 2: Revisión de Álgebra Lineal)

2016/17

Universidad
Carlos III de Madrid

Contenido

Revisión de Álgebra Lineal

Vectores

- **Convención**:

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}$$

  es un **vector columna** de $p \times 1$.

- Por tanto,

$$\mathbf{a}^T = \begin{pmatrix} a_1 & a_2 & \cdots & a_p \end{pmatrix}$$

  es un **vector fila** de $1 \times p$.

Revisión de Álgebra Lineal

Matrices

- Una **matrix** de $p \times q$ es un arreglo de la forma

$$\mathbf{A} = (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{pmatrix}.$$

- Expresiones en **filas** y **columnas** ($\mathbf{a}_i^T$ de $1 \times q$ y $\boldsymbol{\alpha}_j$ de $p \times 1$):

$$\mathbf{A} = (a_{ij}) = \begin{pmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_p^T \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}_1 & \boldsymbol{\alpha}_2 & \cdots & \boldsymbol{\alpha}_q \end{pmatrix}.$$

Revisión de Álgebra Lineal

Ejemplo

- Para las matrices

$$\mathbf{A} = \begin{pmatrix} 4 & 2 & 7 \\ 1 & 8 & 3 \end{pmatrix} \quad , \quad \mathbf{B} = \begin{pmatrix} 1 & 2 \\ -1 & -3 \\ 0 & 1 \end{pmatrix} :$$

- $\mathbf{a}_1^T = \begin{pmatrix} 4 & 2 & 7 \end{pmatrix}$, $\mathbf{a}_2^T = \begin{pmatrix} 1 & 8 & 3 \end{pmatrix}$.
- $\boldsymbol{\alpha}_1 = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$, $\boldsymbol{\alpha}_2 = \begin{pmatrix} 2 \\ 8 \end{pmatrix}$, $\boldsymbol{\alpha}_3 = \begin{pmatrix} 7 \\ 3 \end{pmatrix}$.
- $\mathbf{b}_1^T = \begin{pmatrix} 1 & 2 \end{pmatrix}$, $\boldsymbol{\beta}_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \ldots.$

Revisión de Álgebra Lineal

Propiedades

Operaciones usuales (para vectores y matrices de dimensiones adecuadas):

- (**Producto escalar**) $\mathbf{a}^T\mathbf{b}$.
- (**Suma**) $\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$. $\mathbf{A}+(\mathbf{B}+\mathbf{C})=(\mathbf{A}+\mathbf{B})+\mathbf{C}$.
- (**Multiplicación por un escalar**) $c(\mathbf{AB})=(c\mathbf{A})\mathbf{B}=\mathbf{A}(c\mathbf{B})$.
- (**Distributivas**) $\mathbf{A}(\mathbf{B}+\mathbf{C})=\mathbf{AB}+\mathbf{AC}$. $(\mathbf{A}+\mathbf{B})\mathbf{C}=\mathbf{AC}+\mathbf{BC}$.
- (**Determinante**) $|\mathbf{A}|$.
- (**Inversa**) $\mathbf{AB}=\mathbf{BA}=\mathbf{I}_p$, $\mathbf{B}=\mathbf{A}^{-1}$ $(|\mathbf{A}|\neq 0)$.

Revisión de Álgebra Lineal

Traspuesta

Para $\mathbf{A} = (a_{ij})$ de $p \times q$:

$$\text{(Traspuesta)} \quad \mathbf{A}^T = (a_{ji}) \quad p \times q \to q \times p \ .$$

- (**Propiedades**)

$$(\mathbf{A}^T)^T = \mathbf{A}. \quad (\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T. \quad (\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T.$$

- (**Simetría**) $\quad \mathbf{A}^T = \mathbf{A}.$

- (**Filas y columnas**) $^{(\star)}$ Si $\mathbf{A} = (\mathbf{a}_i^T) = (\boldsymbol{\alpha}_j)$:

$$\mathbf{A}^T = (\boldsymbol{\alpha}_j^T) = (\mathbf{a}_i) \ .$$

Revisión de Álgebra Lineal

Ejemplo

- Para la matriz
$$\mathbf{A} = \begin{pmatrix} 4 & 2 & 7 \\ 1 & 8 & 3 \end{pmatrix}:$$
- $\mathbf{a}_1^T = \begin{pmatrix} 4 & 2 & 7 \end{pmatrix}$, $\mathbf{a}_2^T = \begin{pmatrix} 1 & 8 & 3 \end{pmatrix}$.
- $\alpha_1 = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$, $\alpha_2 = \begin{pmatrix} 2 \\ 8 \end{pmatrix}$, $\alpha_3 = \begin{pmatrix} 7 \\ 3 \end{pmatrix}$.
- Entonces:
$$\mathbf{A}^T = \begin{pmatrix} 4 & 1 \\ 2 & 8 \\ 7 & 3 \end{pmatrix} = \begin{pmatrix} \alpha_1^T \\ \alpha_2^T \\ \alpha_3^T \end{pmatrix} = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{pmatrix}.$$

Revisión de Álgebra Lineal

Producto

Para $\mathbf{A} = (a_{ij})$ de $p \times q$ y $\mathbf{B} = (b_{ij})$ de $q \times r$:

- (**Filas**$\times$**Columnas**) Si $\mathbf{A} = (\mathbf{a}_i^T)$ y $\mathbf{B} = (\beta_j)$:

$$\mathbf{AB} = (\mathbf{a}_i^T \beta_j) \qquad (p \times \underbrace{q)(q}_{\text{internas}} \times r) \rightarrow \underbrace{p \times r}_{\text{externas}}.$$

- (**Columnas**$\times$**Filas**) $^{(\star)}$ Si $\mathbf{A} = (\alpha_k)$ y $\mathbf{B} = (\mathbf{b}_k^T)$:

$$\mathbf{AB} = \alpha_1 \mathbf{b}_1^T + \cdots + \alpha_q \mathbf{b}_q^T \qquad (p \times \underbrace{1)(1}_{\text{internas}} \times r) \rightarrow \underbrace{p \times r}_{\text{externas}}.$$

Revisión de Álgebra Lineal

Ejemplo

- Para las matrices

$$\mathbf{A} = \begin{pmatrix} 4 & 2 & 7 \\ 1 & 8 & 3 \end{pmatrix} \quad , \quad \mathbf{B} = \begin{pmatrix} 1 & 2 \\ -1 & -3 \\ 0 & 1 \end{pmatrix} :$$

- $\mathbf{AB} = (\mathbf{a}_i^T \beta_j) = \begin{pmatrix} 2 & 9 \\ -7 & -19 \end{pmatrix}.$

- $\mathbf{AB} = \alpha_1 \mathbf{b}_1^T + \alpha_2 \mathbf{b}_2^T + \alpha_3 \mathbf{b}_3^T =$

$$= \begin{pmatrix} 4 & 8 \\ 1 & 2 \end{pmatrix} + \begin{pmatrix} -2 & -6 \\ -8 & -24 \end{pmatrix} + \begin{pmatrix} 0 & 7 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} 2 & 9 \\ -7 & -19 \end{pmatrix} .$$

Revisión de Álgebra Lineal

Traza

La **traza** de $\mathbf{A} = (a_{ij})$ de $p \times p$ es:

$$\mathrm{tr}(\mathbf{A}) = a_{11} + a_{22} + \cdots + a_{pp} = \sum_{i=1}^{p} a_{ii} \ .$$

Teorema

- $\mathrm{tr}(c\mathbf{A}) = c\,\mathrm{tr}(\mathbf{A})$.
- $\mathrm{tr}(\mathbf{A} + \mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$.
- $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$ $^{(\star)}$.

- $^{(\star)}$ En **general**, $\mathbf{AB} \neq \mathbf{BA}$ (quizá **ni** siquiera **definido**).
- Pero si $\mathbf{AB}$ y $\mathbf{BA}$ **existen**, las trazas **coinciden**.

Revisión de Álgebra Lineal

Ejemplo

- Para las matrices

$$\mathbf{A} = \begin{pmatrix} 4 & 2 & 7 \\ 1 & 8 & 3 \end{pmatrix} \quad , \quad \mathbf{B} = \begin{pmatrix} 1 & 2 \\ -1 & -3 \\ 0 & 1 \end{pmatrix} :$$

- $\mathbf{AB} = \begin{pmatrix} 2 & 9 \\ -7 & -19 \end{pmatrix}$. $\mathrm{tr}(\mathbf{AB}) = 2 - 19 = -17$.

- $\mathbf{BA} = \begin{pmatrix} 6 & 18 & 13 \\ -7 & -26 & -16 \\ 1 & 8 & 3 \end{pmatrix}$. $\mathrm{tr}(\mathbf{AB}) = 6 - 26 + 3 = -17$.

Resumen

1. Revisión de Álgebra Lineal

- **Referencias**: Johnson, R.A. y Wichern, D.W. (2007) [Cap. 2].

Análisis de Datos Multivariantes

1. INTRODUCCIÓN

(Parte 3: Combinaciones Lineales)

2016/17

Universidad Carlos III de Madrid

Contenido

Universidad Carlos III de Madrid

Combinaciones Lineales

Universidad Carlos III de Madrid

Combinaciones Lineales

Dadas las variables $\mathbf{X} = (X_1, \ldots, X_p)^T$, una **combinación lineal** es una relación de la forma

$$Y = a_1 X_1 + a_2 X_2 + \cdots + a_p X_p + b \ ,$$

para ciertas constantes b, y $a_1, \ldots, a_p$.

- a_j's: **Coeficientes**. b: **Término independiente**.
- (**Expresión vectorial**) Una combinación lineal se puede escribir

  $$Y = \mathbf{a}^T \mathbf{X} + b = \mathbf{a}^T (\mathbf{X} - \mathbf{c}) \ ,$$

  donde $\mathbf{a} = (a_1, \ldots, a_p)^T$, y $b = -\mathbf{a}^T \mathbf{c}$.
- Usualmente, se consideran **conjuntos** de combinaciones lineales

  $$(Y_1, \ldots, Y_q)^T \ .$$

Combinaciones Lineales

Ejemplo

- $(p = 3 \to q = 2)$

$$\begin{aligned} Y_1 &= 2X_1 + X_2 - X_3 + 1; \\ Y_2 &= X_1 + 3X_2 + 4X_3 + 3. \end{aligned}$$

- **Expresión matricial:**

$$\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}}_{\mathbf{Y}} = \underbrace{\begin{pmatrix} 2 & 1 & -1 \\ 1 & 3 & 4 \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}}_{\mathbf{X}} + \underbrace{\begin{pmatrix} 1 \\ 3 \end{pmatrix}}_{\mathbf{b}}$$

- $\mathbf{Y}$ es de $q \times 1$, $\mathbf{A}$ de $q \times p$, $\mathbf{X}$ de $p \times 1$, y $\mathbf{b}$ de $q \times 1$.

1. Combinaciones Lineales
   - Definición
   - Ejemplo

2. Efectos en Información Muestral
   - Matriz de Datos, Medias y Covarianzas

3. Aplicaciones
   - Centrado
   - Tipificación

Efectos en Información Muestral

Universidad Carlos III de Madrid

Teorema

Dados $\mathbf{A}$ de $q \times p$, $\mathbf{b}$ de $q \times 1$, y una **transformación lineal** de la forma

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b} :$$

- **Matriz de Datos**:

$$\mathcal{X} \to \mathcal{Y} = \mathcal{X}\mathbf{A}^T + \mathbf{1}_n \mathbf{b}^T .$$

- **Vector de medias muestrales**:

$$\bar{\mathbf{x}} \to \bar{\mathbf{y}} = \mathbf{A}\bar{\mathbf{x}} + \mathbf{b} .$$

- **Matriz de covarianzas muestrales**:

$$\mathbf{S} \to \mathbf{S}_{\mathcal{Y}} = \mathbf{A}\mathbf{S}\mathbf{A}^T .$$

- **Dimensiones congruentes**.

1. Combinaciones Lineales
   - Definición
   - Ejemplo

2. Efectos en Información Muestral
   - Matriz de Datos, Medias y Covarianzas

3. Aplicaciones
   - Centrado
   - Tipificación

Aplicaciones

Centrado

- **Datos Centrados**:

$$\mathcal{X} = (x_{ij}) \longrightarrow \mathcal{X}_c = (x_{ij} - \overline{x}_j) \ .$$

- **Transformación genérica**: $\mathbf{Y} = \mathbf{X} - \overline{\mathbf{x}} = \mathbf{A}\mathbf{X} + \mathbf{b}$, $\mathbf{A} = \mathbf{I}_p$, $\mathbf{b} = -\overline{\mathbf{x}}$.

Lema

- **Matriz de Datos**: $\mathcal{X}_c = \mathcal{X}\mathbf{A}^T + \mathbf{1}_n\mathbf{b}^T = \mathcal{X} - \mathbf{1}_n\overline{\mathbf{x}}^T$.
- **Vector de Medias Muestrales** $^{(\star)}$: $\overline{\mathbf{x}}_c = \mathbf{A}\overline{\mathbf{x}} + \mathbf{b} = \overline{\mathbf{x}} - \overline{\mathbf{x}} = \mathbf{0}$.
- **Matriz de Covarianzas Muestrales**: $\mathbf{S}_c = \mathbf{A}\mathbf{S}\mathbf{A}^T = \mathbf{S}$. $(\mathbf{R}_c = \mathbf{R})$.

Aplicaciones

Universidad Carlos III de Madrid

Tipificación

- **Datos Tipificados**: Si $\mathbf{D} = \mathrm{diag}(\mathbf{S})$,

$$\mathcal{X} = (x_{ij}) \longrightarrow \mathcal{X}_s = \left(\frac{x_{ij} - \overline{x}_j}{\sqrt{s_{jj}}} \right) .$$

- **Transformación genérica**:

$$\mathbf{Y} = \mathbf{D}^{-1/2}(\mathbf{X} - \overline{\mathbf{x}}) = \mathbf{A}\mathbf{X} + \mathbf{b}, \quad \mathbf{A} = \mathbf{D}^{-1/2}, \ \mathbf{b} = -\mathbf{D}^{-1/2}\overline{\mathbf{x}} .$$

Lema

- **Datos**: $\mathcal{X}_s = \mathcal{X}\mathbf{A}^T + \mathbf{1}_n\mathbf{b}^T = (\mathcal{X} - \mathbf{1}_n\overline{\mathbf{x}}^T)\mathbf{D}^{-1/2} = \mathcal{X}_c\mathbf{D}^{-1/2}$.
- **Medias muestrales** $^{(\star)}$: $\overline{\mathbf{x}}_s = \mathbf{A}\overline{\mathbf{x}} + \mathbf{b} = \mathbf{D}^{-1/2}(\overline{\mathbf{x}} - \overline{\mathbf{x}}) = \mathbf{0}$.
- **Covarianzas muestrales** $^{(\star\star)}$: $\mathbf{S}_s = \mathbf{A}\mathbf{S}\mathbf{A}^T = \mathbf{D}^{-1/2}\mathbf{S}\mathbf{D}^{-1/2} = \mathbf{R}$.

Aplicaciones$^{(\star)}$

Universidad Carlos III de Madrid

Forma esférica

- **Datos con Forma Esférica**:

$$\mathcal{X} = (x_{ij}) \longrightarrow \mathcal{X}_{sp} = (x_{ij} - \overline{x}_j)\,\mathbf{S}^{-1/2} \; .$$

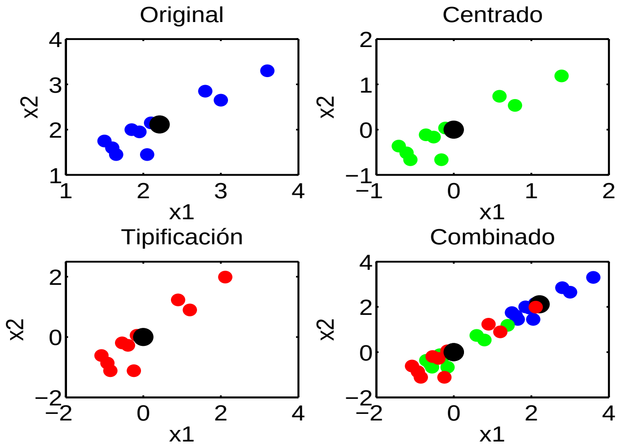
- **Transformación genérica**:

$$\mathbf{Y} = \mathbf{S}^{-1/2}(\mathbf{X} - \overline{\mathbf{x}}) = \mathbf{A}\mathbf{X} + \mathbf{b}, \quad \mathbf{A} = \mathbf{S}^{-1/2},\, \mathbf{b} = -\mathbf{S}^{-1/2}\overline{\mathbf{x}} \; .$$

Lema

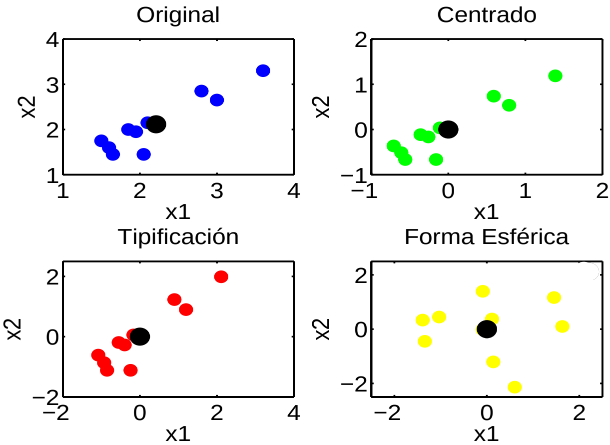
- **Datos**: $\mathcal{X}_{sp} = \mathcal{X}\mathbf{A}^T + \mathbf{1}_n\mathbf{b}^T = (\mathcal{X} - \mathbf{1}_n\overline{\mathbf{x}}^T)\mathbf{S}^{-1/2} = \mathcal{X}_c\mathbf{S}^{-1/2}$.
- **Medias muestrales** $^{(\star)}$: $\overline{\mathbf{x}}_{sp} = \mathbf{A}\overline{\mathbf{x}} + \mathbf{b} = \mathbf{S}^{-1/2}(\overline{\mathbf{x}} - \overline{\mathbf{x}}) = \mathbf{0}$.
- **Covarianzas muestrales** $^{(\star\star)}$: $\mathbf{S}_{sp} = \mathbf{A}\mathbf{S}\mathbf{A}^T = \mathbf{S}^{-1/2}\mathbf{S}\mathbf{S}^{-1/2} = \mathbf{I}_p$.

Aplicaciones



Centrado y Tipificación (Datos de Empresas)

Aplicaciones



Centrado, Tipificación, y Forma Esférica (Datos de Empresas)

Resumen

Universidad Carlos III de Madrid

1. Combinaciones Lineales
2. Efectos en Información Muestral
3. Aplicaciones

- **Referencias**: Johnson, R.A. y Wichern, D.W. (2007) [Cap. 3].