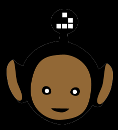
# Clase 5 graficar variables en espacios n dimensionales

Rodrigo Castillo

12 de agosto de 2020

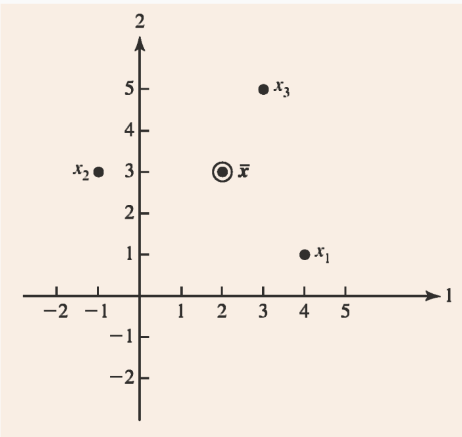


## 1. Ejemplo facil de entender

**Ejemplo**

Graficar el vector media para los siguientes datos tomando $n = 3$ puntos en un espacio de $p = 2$ dimensiones.

$$\mathbf{X} = \begin{bmatrix} 4 & 1 \\ -1 & 3 \\ 3 & 5 \end{bmatrix}$$




el caso anterior se puede extender para N dimensiones

## 2. graficar p vectores en n dimensiones

### 2.1. primer ejemplo

**$p$ vectores en el espacio $n$-dimensional**

- La idea es tomar los elementos de las columnas de la matriz como las coordenadas de los vectores:

$$\underset{(n\times p)}{\mathbf{X}} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1 \mid \mathbf{y}_2 \mid \cdots \mid \mathbf{y}_p \end{bmatrix}$$

- El i-ésimo punto $\mathbf{y}_i' = \begin{bmatrix} x_{1i}, x_{2i}, \ldots, x_{ni} \end{bmatrix}$ está determinado por la

igual que el ejemplo anterior de todas formas

## 3. relacion entre la desviacion y la desviacion estandar

**Relación entre la desviación y la desviación estándar**

$$L_{\mathbf{d}_i}^2 = \mathbf{d}_i'\mathbf{d}_i = \sum_{j=1}^{n} (x_{ji} - \bar{x}_i)^2$$

$$\mathbf{d}_i'\mathbf{d}_k = \sum_{j=1}^{n} (x_{ji} - \bar{x}_i)(x_{jk} - \bar{x}_k)$$

$$\sum_{j=1}^{n} (x_{ji} - \bar{x}_i)(x_{jk} - \bar{x}_k) = \sqrt{\sum_{j=1}^{n} (x_{ji} - \bar{x}_i)^2}\sqrt{\sum_{j=1}^{n} (x_{jk} - \bar{x}_k)^2}\cos(\theta_{ik})$$

$$\cos\theta = \frac{\mathbf{x}'\mathbf{y}}{L_{\mathbf{x}}L_{\mathbf{y}}}$$

**El coseno del ángulo es el coeficiente de correlación**

$$r_{ik} = \frac{s_{ik}}{\sqrt{s_{ii}}\sqrt{s_{kk}}} = \cos(\theta_{ik})$$

$$\mathbf{d}_i'\mathbf{d}_k = L_{\mathbf{d}_i}L_{\mathbf{d}_k}\cos(\theta_{ik})$$

## 4. media poblacional y varianza poblacional

$E(\overline{\mathbf{X}}) = \boldsymbol{\mu}$ (vector de media poblacional)

$\operatorname{Cov}(\overline{\mathbf{X}}) = \frac{1}{n}\boldsymbol{\Sigma}$ (matriz de covarianza poblacional dividida por el tamaño de la muestra)

## 5. estimador para $\sum$

**Estimador para $\boldsymbol{\Sigma}$**

Para la matriz de covarianza $\mathbf{S}_n$,

$$E(\mathbf{S}_n) = \frac{n-1}{n}\boldsymbol{\Sigma} = \boldsymbol{\Sigma} - \frac{1}{n}\boldsymbol{\Sigma} \quad \text{o,} \quad E\left(\frac{n}{n-1}\mathbf{S}_n\right) = \boldsymbol{\Sigma}$$

Entonces, $[n/(n-1)]\mathbf{S}_n$ es un estimador insesgado de $\boldsymbol{\Sigma}$, mientras que $\mathbf{S}_n$ es un estimador sesgado con $sesgo = E(\mathbf{S}_n) - \boldsymbol{\Sigma} = -(1/n)\boldsymbol{\Sigma}$

**Matriz de covarianza muestral insesgada**

$$\mathbf{S} = \left(\frac{n}{n-1}\right)\mathbf{S}_n = \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{X}_j - \overline{\mathbf{X}})(\mathbf{X}_j - \overline{\mathbf{X}})'$$

Puesto que $\mathbf{S}$ se usa generalmente para pruebas multivariables, $\mathbf{S}_n$ será sustituida por $\mathbf{S}$ .

## 6. Preguntas

Preguntar bien como se calcula el estimador para $\sum$ y para que sirve