

# Notas Tema 2 AED

Rodrigo Castillo

5 de septiembre de 2020



## 1. Distancia Euclideana

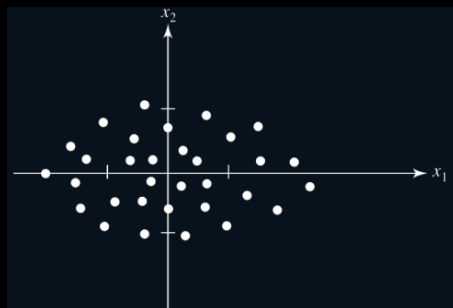
- la mayoría de las técnicas multivariantes dependen del concepto de distancia

sean  $P = (x_1, x_2, x_3, \dots, x_p)$  y  $Q = (y_1, y_2, y_3, \dots, y_p)$  se tiene que la distancia entre P y Q está dada como :

$$d(P, Q) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2} \quad (1)$$

- la distancia euclideana sirve para medir distancias entre 2 puntos pero en estadística no sirve, se usa una distancia llamada Distancia Estadística
- no funciona porque cada coordenada pesa lo mismo
- en la estadística es bueno darle mayor peso a las variables que varían menos y menor peso a aquellas que varían más
- la distancia estadística tiene en cuenta tanto el peso de la variación como el peso de las variables

## 2. Distancia Estadística



### Suposiciones

- Hay  $n$  pares de medidas sobre dos variables  $x_1$  y  $x_2$ .
- Tiene media cero y varían independientemente.
- La variabilidad en  $x_1$  es mayor que en  $x_2$ .

- los puntos respecto al origen en la dirección de  $x_1$  son más comunes que los puntos a la misma distancia en la dirección  $x_2$
- por lo tanto la variabilidad de  $x_1$  es mayor que la de  $x_2$

- son mas comunes coordenadas grandes en  $x_1$  que en  $x_2$
- tiene sentido ponderar mas a  $x_2$  que a  $x_1$

para construir la distancia estadística ...

1.  $x_1^* = \frac{x}{\sqrt{s_{11}}}$  y  $x_2^* = \frac{x_2}{\sqrt{s_{22}}}$  para ponderar las coordenadas con menor variabilidad
2. aplicamos la distancia euclidea para definir la distancia entre el punto  $P = (x_1, x_2)$  al origen  $O = (0, 0)$
3. si la variabilidad de ambas distancias es igual, se usa la distancia euclidea

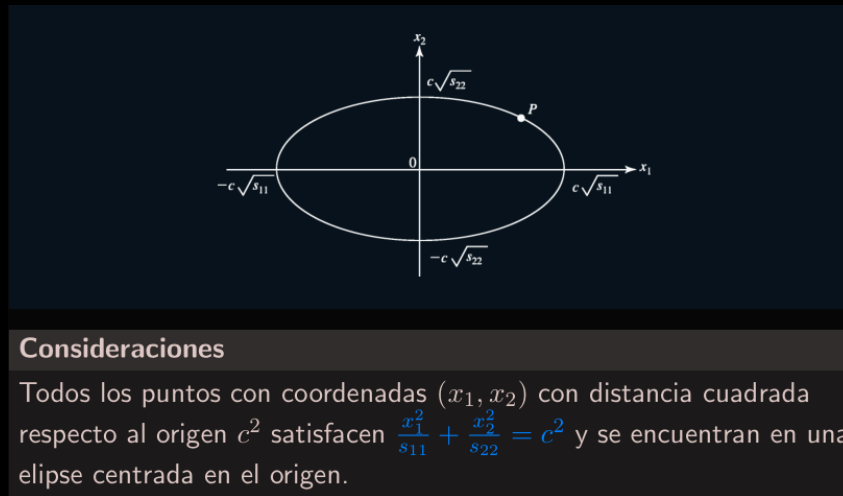


Figura 1: Consideraciones

por lo tanto la distancia estadística está dada como :

$$D_{estadística}(P, Q) = \sqrt{\frac{(x_1 - y_1)^2}{s_{11}} + \frac{(x_2 - y_2)^2}{s_{22}} + \dots + \frac{(x_{pp} - y_p)^2}{s_{pp}}} \quad (2)$$

- todos los puntos  $P$  están a una distancia cuadrada constante del punto  $Q$  yacen sobre una hiper elipsoide centrada en  $Q$  con ejes mayor y menor paralelos a los ejes coordenados
- la distancia desde  $P$  hasta el origen  $O$  se puede obtener mediante  $y_1 = y_2 = y_3 \dots y_p = 0$
- si  $s_{11} = s_{22}$  da igual que usar la distancia euclidea
- la distancia estadística supone que las coordenadas son independientes
- no incluye los casos importantes

### Un caso de variabilidad correlacionada

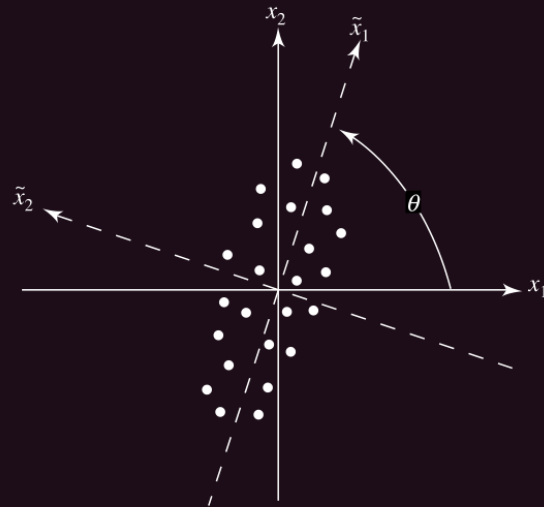


Figura 2: Caso

### Un caso de variabilidad correlacionada

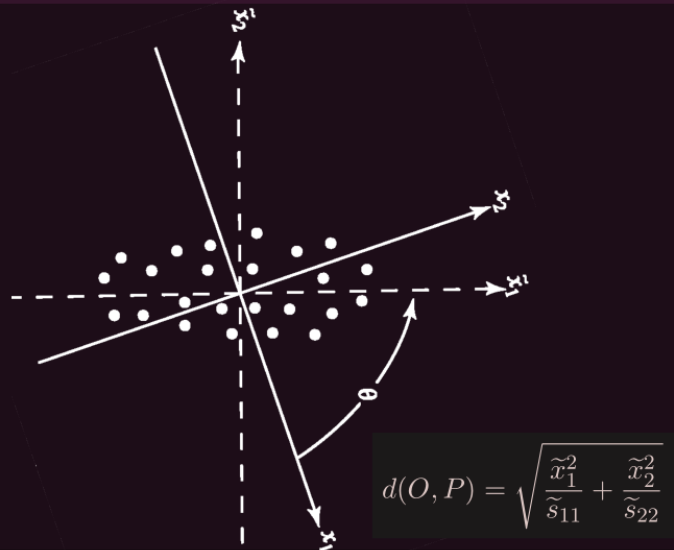


Figura 3: Caso2

cuando las variables están correlacionadas , la variable a un punto **fijo** es ...

$$D(P, Q) = \sqrt{a_{11}(x_1 - y_1)^2 + 2a_{12}(x_1 - y_1)(x_2 - y_2)} \quad (3)$$

esta distancia puede ser calculada una vez se calculen  $a_{11}, a_{12}, a_{22}$

#### Extendiendo la distancia estadística a $p$ dimensiones

- Sea  $P = (x_1, x_2, \dots, x_p)$  un punto cuyas coordenadas representan variables que están correlacionadas y sujetas a variabilidad.

- Sea  $O = (0, 0, \dots, 0)$  el origen.

- Sea  $Q = (y_1, y_2, \dots, y_p)$  un punto fijo especificado.

- La distancia desde  $P$  a  $O$  es:  $d(O, P) = \sqrt{a_{11}x_1^2 + a_{22}x_2^2 + \dots + a_{pp}x_p^2 + 2a_{12}x_1x_2 + 2a_{13}x_1x_3 + \dots + 2a_{p-1,p}x_{p-1}x_p}$

- La distancia desde  $P$  a  $Q$ ,  $d(P, Q)$  es:

$$\sqrt{[a_{11}(x_1 - y_1)^2 + a_{22}(x_2 - y_2)^2 + \dots + a_{pp}(x_p - y_p)^2 + 2a_{12}(x_1 - y_1)(x_2 - y_2) + 2a_{13}(x_1 - y_1)(x_3 - y_3) + \dots + 2a_{p-1,p}(x_{p-1} - y_{p-1})(x_p - y_p)]}$$

Figura 4: P-Dimensiones

Ahora, de forma matricial la distancia queda como :

- Los coeficientes (pesos)  $a_{ik}$  determinan completamente la distancia.
- Los coeficientes se pueden reorganizar de forma matricial:

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{12} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1p} & a_{2p} & \dots & a_{pp} \end{bmatrix}$$

Figura 5: Matricial

### 3. cosas de repaso

producto interno

$$x'y = x_1y_1 + x_2y_2 + \dots + x_ny_n \quad (4)$$

longitud de un vector

$$L_x = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} = \sqrt{x'x} \quad (5)$$

angulo entre dos vectores ...

$$\cos \theta = \frac{x'y}{L_x L_y} = \frac{x'y}{\sqrt{x'x} \sqrt{y'y}} \quad (6)$$

independencia lineal

un conjunto de vectores  $x_1, x_2, x_3, \dots, x_n$  si existen constantes  $c_1 \dots c_n$  tales que  $x_1c_1 + c_2x_2 + \dots + c_nx_n = 0$

### 4. Dudas

1. no me queda claro la ecuacion 2 de donde sale la ecuacion 3