

PROBLEMAS TEMA 4: MEDIDAS DE PROXIMIDAD

Problema 1 Considera los puntos $A = (1, 2)$, $B = (2, 1)$ y $C = (-2, 1)$.

- (a) Utilizando como función de distancia la distancia euclídea al cuadrado, comprueba que se cumple la desigualdad triangular

$$d_e^2(A, B) + d_e^2(C, B) \geq d_e^2(A, C),$$

- (b) Idem, pero ahora usando la distancia L_1 o de Manhattan.

Problema 2 Se han medido las siguientes variables a los 5 modelos de jersey de la misma marca:

Modelo	Precio	Gasto en publicidad	Presencia (horas) de un asistente de ventas
A	230	125	200
B	181	99	55
C	165	97	105
D	150	115	85
E	97	120	0

- (a) Calcular la matriz de distancias euclídeas al cuadrado entre los 5 modelos de jersey. ¿Qué par de modelos de jersey se parecen más? ¿Qué par se parecen menos?
- (b) ¿Te parecen razonables las conclusiones obtenidas en (a)? ¿Qué otra/s distancia/s sería más conveniente usar?

Problema 3 La siguiente tabla de contingencia representa la distribución conjunta de 230 personas respecto a las variables *Estado Civil* y *Sexo*.

	Mujer	Hombre	Total
Soltero	112	8	120
Casado	107	3	110
Total	219	11	230

La *tabla de perfiles columna* se calcula dividiendo cada elemento de la tabla entre el total de su columna. Las columnas de esta tabla se llaman *puntos columna* y la última columna contiene las frecuencias relativas marginales de las filas. Calcula la distancia χ^2 entre los puntos columna, tomando como pesos los inversos de las frecuencias relativas marginales de las filas.

Problema 4 (*Examen Feb. 2004*) En la tabla adjunta aparecen ciertas características medidas a tres presidentes de Estados Unidos.

Presidente	Lugar nacimiento	Elegido en la primera vuelta	Partido	Experiencia anterior en el congreso	Vicepresidente
Reagan	Medio Oeste	Sí	Republicano	No	No
Ford	Medio Oeste	No	Republicano	Sí	Sí
Kennedy	Este	Sí	Demócrata	Sí	No

A partir de los datos de la tabla, obtener una matriz de similitudes entre los tres presidentes.

Problema 5 (*Examen Sept. 2004*) Se desea averiguar cómo perciben los televidentes las noticias de distintos canales de televisión. Para ello, se seleccionan cuatro canales de televisión nacionales, A , B , C y D . Se hacen todos los pares posibles de canales de televisión (hay 6 posibles pares), y se le pide a un televidente que asigne un rango de 1 a 6 a dichos pares, de manera que los canales del par con rango 1 sean los que le parecen más indiferentes, mientras que los canales del par con un rango 6 sean los que más diferencias presentan en cuanto a sus preferencias. Este televidente nos da la siguiente ordenación

Par	(C,D)	(B,C)	(B,D)	(A,D)	(A,B)	(A,C)
Rango	1	2	3	4	5	6

Esta ordenación indica que, para el televidente en cuestión, los canales más indiferentes son C y D , mientras que los menos indiferentes son A y C . A partir de estos rangos, obtener una matriz de disimilitudes entre los distintos canales de tv de la siguiente forma: para cada canal de tv, calcula su rango medio promediando los rangos de los distintos pares en los que aparece. La disimilitud final entre dos canales de tv va a ser la diferencia en valor absoluto entre sus rangos medios.

Problema 6 (*Examen Feb. 2008*) Se han tomado datos sobre 6 jugadores de fútbol de la Primera División de la liga española 2006/2007. las variables medidas son:

X_1 : Pierna buena del jugador (0=derecha, 1=izquierda)

X_2 : Nacionalidad (1=Latinoamericano, 2=Africano, 3=Europeo)

X_3 : Nivel de estudios (1=Sin estudios, 2=Estudios Primarios, 3=Estudios secundarios)

Los datos obtenidos son

Jugador	X_1	X_2	X_3
1. Ronaldinho	1	1	2
2. Etoo	0	2	2
3. Messi	0	1	3
4. Raúl	1	3	3
5. Ronaldo	0	1	1
6. Beckham	0	3	3

- (a) Utilizando el coeficiente de comparación simple y teniendo en cuenta que la variable X_3 (Nivel de estudios) es ORDINAL, obtener una matriz de similitudes $S = (s_{ij})$ entre los jugadores.
- (b) Calcula una matriz de disimilitudes entre jugadores, $\Delta = (\delta_{ij})$, donde la disimilitud entre los jugadores i y j se define por

$$\delta_{ij} = 9(1 - s_{ij}).$$

Problema 7 (*Examen Feb. 2009*) Se dispone de las siguientes variables para 5 clientes de un banco:

X_1 : Antigüedad del cliente (1=<1 año, 2=entre 1 y 5 años, 3=>5 años)

X_2 : Tipo de contrato laboral (1=Temporal, 2=Indefinido)

X_3 : Posee propiedades inmobiliarias (1=Sí, 2=No)

Los datos obtenidos son:

Cliente	X_1	X_2	X_3
1	1	1	2
2	2	2	2
3	3	1	1
4	2	2	1
5	3	1	2

- (a) Utilizando el coeficiente de comparación simple, calcular una matriz de similitudes $S = (s_{ij})$ entre los clientes, teniendo en cuenta que la variable X_1 (Antigüedad) es ORDINAL.
- (b) Calcula una matriz de disimilitudes entre clientes, $\Delta = (\delta_{ij})$, donde la disimilitud entre los clientes i y j se define por

$$\delta_{ij} = 6(1 - s_{ij}).$$