Godinez Bravo Diego

Estadística Multivariada

Tarea 5

CIMAT

Ejercicio 1. En muchas situaciones las variables que se observan sobre un conjunto de individuos son de naturaleza binaria. En estos casos para poder disponer de una matriz de distancias entre individuos se utilizan coeficientes de similaridad. El coeficiente de similaridad entre el individuo i y el individuo j, s_{ij}, se calcula a partir de las frecuencias:

$a =$ "número de variables con respuesta 1 en ambos individuos"

$b =$ "número de variables con respuesta 0 en el primer individuo y con respuesta 1 en el segundo individuo"

$c =$ "número de variables con respuesta 1 en el primer individuo y con respuesta 0 en el segundo individuo"

$d =$ "número de variables con respuesta 0 en ambos individuos"

Existen muchos coeficientes de similaridad, pero los de Sokal-Michener y de Jacard son especialmente interesantes porque dan lugar a una configuración euclidiana. Se definen como:

$$Sokal - Michener : s_{ik} = \frac{a+d}{p}, \qquad Jacard : s_{ik} = \frac{a}{a+b+c}$$

donde p es el número de variables observadas. Aplicando uno de estos coeficientes a un conjunto de n individuos se obtiene una matriz de similaridades $S = [s_{ij}]_{n\times n}$. Utilizando la siguiente transformación podemos convertir la matriz de similaridades a una matriz de distancias:

$$D^2 = 2(1_n 1_n^{'} - S)$$

Se considera el siguiente conjunto de 6 individuos formado por 5 animales, león, jirafa, vaca, oveja, gato doméstico, junto con el hombre. Se miden 6 variables binarias sobre estos individuos: $X_1 =$ tiene cola, $X_2 =$ es salvaje, $X_3 =$ tiene el cuello largo, $X_4 =$ es animal de granja, $X_5 =$ es carnívoro y $X_6 =$ camina sobre 4 patas.

(a) Obtenga la matriz de datos.

(b) Calcule los coeficientes de similaridad de Sokal-Michener y de Jacard para cada par de individuos y obtenga las matrices de distancias asociadas.

Solución

La matriz representa un conjunto de datos que describe la presencia de 6 características en 6 organismos diferentes. Cada fila corresponde a un individuo, mientras que cada columna representa una característica observada en cada uno de los animales. Los valores en la matriz son 1 si el animal posee la característica correspondiente y 0 en caso contrario.

	X_1	X_2	X_3	X_4	X_5	X_6
León	1	1	0	0	1	1
Jirafa	1	1	1	0	0	1
Vaca	1	0	0	1	0	1
Oveja	1	0	0	1	0	1
Gato Doméstico	1	0	0	0	1	1
Humano	0	0	0	0	1	0

Solución

Los coeficientes de similaridad de **Sokal-Michener** se definen como

$$s_{ik} = \frac{a+d}{p}$$

donde a es el número de variables con respuesta 1 en ambos individuos; d el número de variables con respuesta 0 en ambos individuos; y p es el número de variables observadas.

Por lo tanto, se calcularon los coeficientes de similaridad de la siguiente manera

$$s_{Leon,Jirafa} = \frac{3+1}{6} = \frac{2}{3}$$

$$\vdots$$

$$s_{Gato,Humano} = \frac{1+3}{6} = \frac{2}{3}$$

Realizando los cálculos correspondientes obtuvimos los coeficientes de similaridad de **Sokal-Michener**.

	León	Jirafa	Vaca	Oveja	Gato Doméstico	
Jirafa	0,66					
Vaca	0,50	0,50				
Oveja	0,50	0,50	1,00			
Gato Doméstico	0,66	0,33	0,66	0,66		
Humano	0,50	0,16	0,33	0,33		0,66

La matriz de distancias asociada se puede obtener a partir de la matriz de similaridades considerando la siguiente relación

$$D^2 = 2(1_n 1_n^{'} - S)$$

Sea la matriz de similaridades $S = [s_{ij}]_{n\times n}$ obtenida

$$S = \begin{pmatrix} 1{,}00 & & & & & \\ 0{,}66 & 1{,}00 & & & & \\ 0{,}50 & 0{,}50 & 1{,}00 & & & \\ 0{,}50 & 0{,}50 & 1{,}00 & 1{,}00 & & \\ 0{,}66 & 0{,}33 & 0{,}66 & 0{,}66 & 1{,}00 & \\ 0{,}50 & 0{,}16 & 0{,}33 & 0{,}33 & 0{,}66 & 1{,}00 \end{pmatrix}$$

y dado que $n = 6$ entonces $1_n 1_n^{'}$ resulta en una matriz de tamaño 6×6 donde cada uno de los elementos de la matriz es 1, por lo tanto

$$1_n 1_n^{'} - S = \begin{pmatrix} 0 & & & & & \\ 0{,}34 & 0 & & & & \\ 0{,}50 & 0{,}50 & 0 & & & \\ 0{,}50 & 0{,}50 & 0 & 0 & & \\ 0{,}34 & 0{,}67 & 0{,}34 & 0{,}34 & 0 & \\ 0{,}50 & 0{,}84 & 0{,}67 & 0{,}67 & 0{,}34 & 0 \end{pmatrix}$$

Por lo tanto la matriz de distancias asociada queda definida de la siguiente manera

$$D^2 = \begin{pmatrix} 0 & & & & & \\ 0{,}68 & 0 & & & & \\ 1{,}00 & 1{,}00 & 0 & & & \\ 1{,}00 & 1{,}00 & 0 & 0 & & \\ 0{,}68 & 1{,}34 & 0{,}68 & 0{,}68 & 0 & \\ 1{,}00 & 1{,}68 & 1{,}34 & 1{,}34 & 0{,}68 & 0 \end{pmatrix}$$

Los coeficientes de similaridad de **Jaccard** se definen como

$$s_{ik} = \frac{a}{a+b+c}$$

donde a es el número de variables con respuesta 1 en ambos individuos; b número de variables con respuesta 0 en el primer individuo y con respuesta 1 en el segundo individuo; c número de variables con respuesta 1 en el primer individuo y con respuesta 0 en el segundo individuo; y p es el número de variables observadas.

De manera que se calcularon los coeficientes de similaridad de la siguiente manera

$$s_{Leon,Jirafa} = \frac{3}{3+1+1} = \frac{3}{5}$$

$$\vdots$$

$$s_{Gato,Humano} = \frac{1}{1+2} = \frac{1}{3}$$

Realizando los cálculos correspondientes obtuvimos los coeficientes de similaridad de **Jaccard**.

	León	Jirafa	Vaca	Oveja	Gato Doméstico
Jirafa	0,60				
Vaca	0,40	0,40			
Oveja	0,40	0,40	1,00		
Gato Doméstico	0,75	0,40	0,50	0,50	
Humano	0,25	0,00	0,00	0,00	0,33

Sea la matriz de distancias definida como

$$D^2 = 2(1_n 1_n^{'} - S)$$

y considerando la matriz de similaridades S obtenida

$$S = \begin{pmatrix} 1{,}00 & & & & & \\ 0{,}60 & 1{,}00 & & & & \\ 0{,}40 & 0{,}40 & 1{,}00 & & & \\ 0{,}40 & 0{,}40 & 1{,}00 & 1{,}00 & & \\ 0{,}75 & 0{,}40 & 0{,}50 & 0{,}50 & 1{,}00 & \\ 0{,}25 & 0{,}00 & 0{,}00 & 0{,}00 & 0{,}33 & 1{,}00 \end{pmatrix}$$

Entonces

$$1_n 1_n^{'} - S = \begin{pmatrix} 0 & & & & & \\ 0{,}40 & 0 & & & & \\ 0{,}60 & 0{,}60 & 0 & & & \\ 0{,}60 & 0{,}60 & 0 & 0 & & \\ 0{,}25 & 0{,}60 & 0{,}50 & 0{,}50 & 0 & \\ 0{,}75 & 1{,}00 & 1{,}00 & 1{,}00 & 0{,}67 & 0 \end{pmatrix}$$

Por lo que la matriz de distancias asociada D^2 queda definida como

$$D^2 = \begin{pmatrix} 0 & & & & & \\ 0{,}80 & 0 & & & & \\ 1{,}20 & 1{,}20 & 000 & & & \\ 1{,}20 & 1{,}20 & 0 & 0 & & \\ 0{,}50 & 1{,}20 & 1{,}00 & 1{,}00 & 0 & \\ 1{,}50 & 2{,}00 & 2{,}00 & 2{,}00 & 1{,}34 & 0 \end{pmatrix}$$

Ejercicio 2. Sea O un conjunto de n individuos cuya matriz de distancias euclidianas es D y cuya representación en coordenadas principales es X. Se desean obtener las coordenadas de un nuevo individuo al que llamaremos individuo $n+1$, del cual se conocen los cuadrados de sus distancias a los n individuos del conjunto O. Si $d = (\delta^2_{n+1,1}, \ldots, \delta^2_{n+1,n})'$ es el vector columna que contiene las distancias al cuadrado del individuo $n+1$ a los restantes n individuos, se puede probar que las coordenadas principales del individuo $n+1$ están dadas por

$$x_{n+1} = \frac{1}{2}\Lambda^{-1}X^{'}(b-d)$$

donde $b = \text{diag}(B) = (b_{11}, \ldots, b_{nn})'$, $B = XX' = U\Lambda U^{'}$ y U es una matriz ortogonal. La ecuación anterior se conoce como la fórmula de interpolación de Gower.

Para los datos del ejercicio 1:

(a) Obtenga una representación en coordenadas principales utilizando la matriz de distancias calculada a partir del coeficiente de similaridad de Sokal-Michener.

(b) Sin volver a recalcular las coordenadas principales, añada el elefante al conjunto de animales y obtenga sus coordenadas principales.

Solución

Haciendo uso del lenguaje de programación R se obtuvo la representación de las coordenadas principales a partir de la matriz de distancias calculada con los coeficientes de similaridad de **Sokal-Michener**.

Basándonos en la representación obtenida podemos concluir que los animales de granja al ser muy similares tienen una distancia nula entre ellos, al igual que los animales salvajes que se encuentran muy cercanos entre sí. Además, observamos una clara separación entre el ser humano y tanto de los animales de granja como los animales salvajes (**Figura 1**).

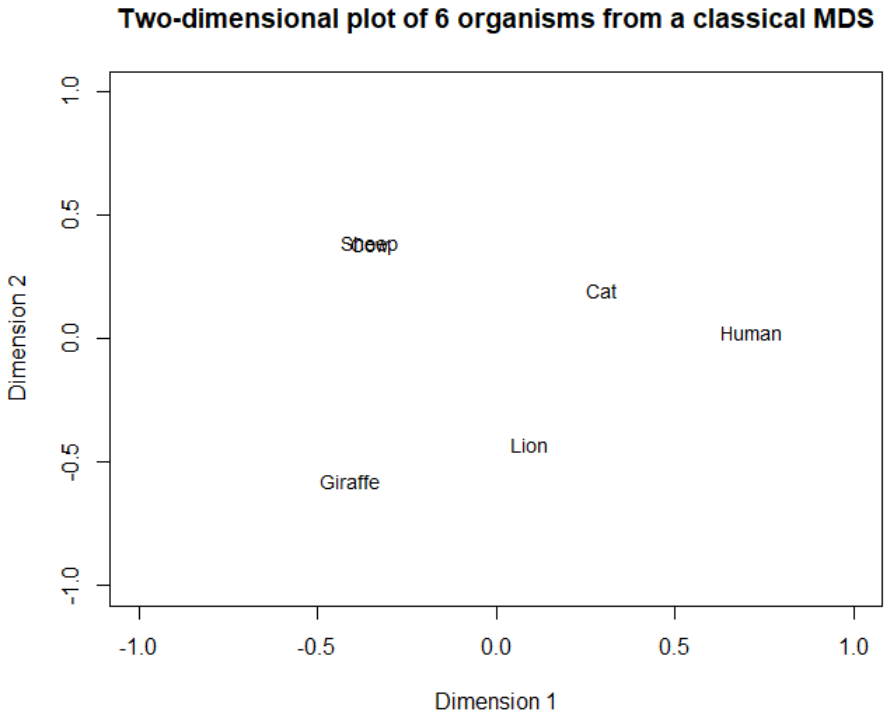

Figura 1. Configuración obtenida mediante el modelo clásico de escalamiento multidimensional.

Solución

Al añadir al elefante al conjunto de individuos la matriz de datos se actualiza de la siguiente manera

	X_1	X_2	X_3	X_4	X_5	X_6
León	1	1	0	0	1	1
Jirafa	1	1	1	0	0	1
Vaca	1	0	0	1	0	1
Oveja	1	0	0	1	0	1
Gato Doméstico	1	0	0	0	1	1
Humano	0	0	0	0	1	0
Elefante	1	1	0	0	0	1

Por lo tanto, la matriz de similaridades $S = [s_{ij}]_{n\times n}$ y la matriz de distancias D^2 se actualizan de la siguiente manera

$$S = \begin{pmatrix} 1{,}00 & & & & & & \\ 0{,}66 & 1{,}00 & & & & & \\ 0{,}50 & 0{,}50 & 1{,}00 & & & & \\ 0{,}50 & 0{,}50 & 1{,}00 & 1{,}00 & & & \\ 0{,}66 & 0{,}33 & 0{,}66 & 0{,}66 & 1{,}00 & & \\ 0{,}50 & 0{,}16 & 0{,}33 & 0{,}33 & 0{,}66 & 1{,}00 & \\ 0{,}83 & 0{,}83 & 0{,}66 & 0{,}66 & 0{,}66 & 0{,}33 & 1{,}00 \end{pmatrix}$$

matriz S a partir de coeficientes de similaridad Sokal-Michener

$$D^2 = \begin{pmatrix} 0 & & & & & & \\ 0{,}68 & 0 & & & & & \\ 1{,}00 & 1{,}00 & 0 & & & & \\ 1{,}00 & 1{,}00 & 0 & 0 & & & \\ 0{,}68 & 1{,}34 & 0{,}68 & 0{,}68 & 0 & & \\ 1{,}00 & 1{,}68 & 1{,}34 & 1{,}34 & 0{,}68 & 0 & \\ 0{,}34 & 0{,}34 & 0{,}68 & 0{,}68 & 0{,}68 & 1{,}34 & 0 \end{pmatrix}$$

matriz de distancias D^2 asociada

Considerando que las coordenadas principales del individuo añadido x_{n+1} están dadas por

$$x_{n+1} = \frac{1}{2}\Lambda^{-1}X^{'}(b-d)$$

y con ayuda del lenguaje de programación R, se calcularon las coordenadas principales del organismo añadido, en este caso, un elefante.

La representación obtenida nos muestra que el nuevo individuo, al compartir características de un animal salvaje se ubica próximo a organismos como el león y la jirafa, con los cuales comparte características fisiológicas, de comportamiento, etc. (**Figura2**).

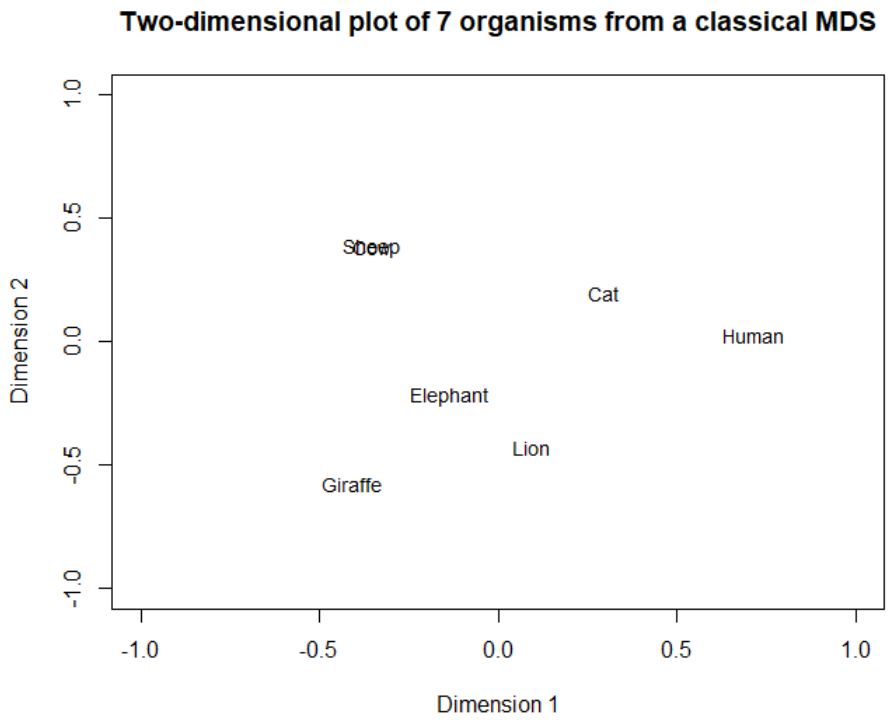

Figura 2. Configuración obtenida mediante el modelo clásico de escalamiento multidimensional.

Ejercicio 3. Una situación muy habitual en el análisis multivariado es disponer de un conjunto de datos mixtos, es decir, un conjunto de individuos sobre los que se han observado tanto variables cuantitativas como cualitativas (o categóricas). En estos casos es de gran utilidad la distancia de Gower, cuyo cuadrado se define como $\delta_{ij}^2 = 1 - s_{ij}$, donde

$$s_{ij} = \frac{\sum_{h=1}^{p_1}(1 - |x_{ih} - x_{jh}|/G_h) + a + \alpha}{p_1 + (p_2 + d) + p_3}$$

es el coeficiente de similaridad de Gower; p_1 es el número de variables cuantitativas continuas, p_2 es el número de variables binarias, p_3 es el número de variables cualitativas (no binarias), a es el número de coincidencias (1,1) en las variables binarias, d es el número de coincidencias (0,0) en las variables binarias, α es el número de coincidencias en las variables cualitativas (no binarias), y G_h es el rango (o recorrido) de la h-ésima variable cuantitativa.

Si $p_1 = p_3 = 0$, entonces s_{ij} de la expresión de arriba coincide con el coeficiente de similaridad de Jaccard. Si se consideran las variables binarias como categóricas (es decir, $p_1 = p_2 = 0$), entonces s_{ij} coincide con el coeficiente de similaridad de Sokal-Michener.

La siguiente tabla contiene la información sobre 50 jugadores de fútbol de la liga española (temporada 2006/2007). Las variables observadas son:

- X_1: número de goles marcados,
- X_2: edad (años),
- X_3: altura (m),
- X_4: peso (kg),
- X_5: pierna buena del jugador (1=derecha, 0=izquierda),
- X_6: Nacionalidad (1=Argentina, 2=Brasil, 3=Camerún, 4=Italia, 5=España, 6=Francia, 7=Uruguay, 8=Portugal, 9=Inglaterra),
- X_7: tipo de estudios (1=sin estudios, 2=básicos, 3=medios, 4=superiores).

(a) Obtenga la matriz de distancias de Gower entre estos jugadores.

(b) Utilizando la matriz de distancias de Gower del inciso anterior, obtenga una representación de los jugadores en coordenadas principales. Determine cuál es el porcentaje de variabilidad explicado por las dos primeras coordenadas principales. ¿Qué puede decir de las semejanzas entre jugadores?

Solución

Haciendo uso del lenguaje de programación R, se calculó la matriz de distancias de Gower utilizando la función *'gower.dist'* la cual forma parte de la librería 'StatMatch'. Se obtuvo una matriz de dimensión 50×50 la cual representa las distancias (disimilaridades) entre los jugadores.

Solución

A partir de la matriz de distancias de Gower se obtuvo una representación a través del modelo clásico de MDS. Observamos una distinción entre ciertos grupos de jugadores que refleja las similitudes y diferencias entre ellos en función de las variables consideradas en el análisis (edad, altura, peso, etc.). Se requiere una investigación adicional para comprender qué características contribuyeron en mayor medida a la separación de ciertos grupos. Es necesario determinar si los jugadores que están cercanos entre sí son similares en cuanto a atributos físicos (edad, altura, peso, etc.), o si el desempeño jugó un papel predominante en la separación de estos grupos. De manera general, esto proporciona una perspectiva visual sobre la diversidad en el conjunto de datos de los jugadores pertenecientes a la liga española (**Figura 3**).

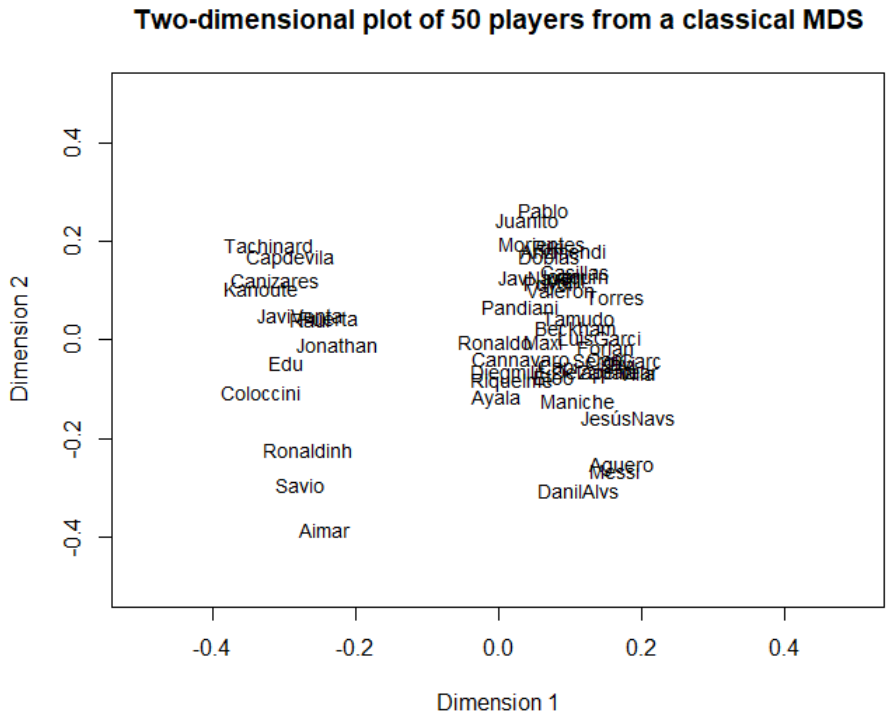

Figura 3. Configuración obtenida mediante el modelo clásico de escalamiento multidimensional a partir de la matriz de distancias de Gower.

Varianza explicada por cada una de las componentes

0,3718 0,3718

Ejercicio 4. Frecuentemente en las aplicaciones nos encontramos con una variable categórica nominal con k estados excluyentes medida sobre una muestra de $n = n_1 + \ldots + n_g$ individuos provenientes de g poblaciones. Se desea obtener una medida de disimilaridad entre estas poblaciones. En estas condiciones, el vector de frecuencias de cada población $n_i = (n_{i1}, \ldots, n_{ik})$, para $i = 1, \ldots, g$, tiene una distribución conjunta multinomial con parámetros (n_i, p_i), donde $n_i = n_{i1} + \ldots + n_{ik}$ y $p_i = (p_{i1}, \ldots, p_{ik})$. Una medida de disimilaridad es la distancia de Bhattacharyya, conocida en genética como distancia Cavalli-Sforza, cuya expresión es:

$$d_{ij}^2 = \arccos\left(\sum_{l=1}^{k} \sqrt{p_{il}p_{jl}}\right)$$

La siguiente tabla contiene las proporciones génicas (observadas) de los grupos sanguíneos correspondientes a 10 poblaciones:

	Población	Grupo A	Grupo AB	Grupo B	Grupo O
1	Francesa	0.21	0.06	0.06	0.67
2	Checa	0.25	0.04	0.14	0.57
3	Germanica	0.22	0.06	0.08	0.64
4	Vasca	0.19	0.04	0.02	0.75
5	China	0.18	0.00	0.15	0.67
6	Ainu	0.23	0.00	0.28	0.49
7	Esquimal	0.30	0.00	0.06	0.64
8	Afroamericana USA	0.10	0.06	0.13	0.71
9	Española	0.27	0.04	0.06	0.63
10	Egipcia	0.21	0.05	0.20	0.54

(a) Obtenga las distancias de Bhattacharyya entre estas poblaciones.

(b) Construye una configuración MDS de las poblaciones mediante la solución clásica (coordenadas principales), utilizando la matriz de distancias Bhattacharyya.

(c) ¿Cuál es la dimensión adecuada de la representación euclidiana? ¿Cuál es el porcentaje de la variabilidad explicada por las dos primeras coordenadas principales? Grafica las poblaciones con las dos primeras coordenadas.

(d) Construye una configuración MDS de las poblaciones utilizando el enfoque de mínimos cuadrados considerando la matriz de distancias Bhattacharyya, tomando como solución inicial la solución clásica y considerando las transformaciones de tipo razón, intervalo y ordinal para las disimilaridades. Compara los resultados obtenidos en cada modelo y justifica la dimensionalidad adecuada de representación y grafica las dos primeras dimensiones.

(e) Compara las configuraciones MDS obtenidas con el enfoque clásico y de mínimos cuadrados. ¿Existen diferencias? ¿Cuáles son las conclusiones?

Solución

Considerando cómo se encuentra definido cada término d_{ij}^2 de la matriz de distancias de Cavalli-Sforza

$$d_{ij}^2 = \arccos\left(\sum_{l=1}^{k}\sqrt{p_{il}p_{jl}}\right)$$

y con apoyo del lenguaje de programación R, se obtuvo el cálculo de la matriz de distancias.

Poblacion	1	2	3	4	5	6	7	8	9	10
1	0,00	0,16	0,04	0,12	0,29	0,40	0,26	0,18	0,08	0,22
2	0,16	0,00	0,12	0,27	0,22	0,26	0,25	0,21	0,14	0,09
3	0,04	0,12	0,00	0,17	0,27	0,36	0,26	0,18	0,08	0,18
4	0,12	0,27	0,17	0,00	0,32	0,47	0,26	0,26	0,15	0,33
5	0,29	0,22	0,27	0,32	0,00	0,19	0,19	0,27	0,27	0,25
6	0,40	0,26	0,36	0,47	0,19	0,00	0,31	0,37	0,36	0,24
7	0,26	0,25	0,26	0,26	0,19	0,31	0,00	0,36	0,20	0,32
8	0,18	0,21	0,18	0,26	0,27	0,37	0,36	0,00	0,24	0,20
9	0,08	0,14	0,08	0,15	0,27	0,36	0,20	0,24	0,00	0,22
10	0,22	0,09	0,18	0,33	0,25	0,24	0,32	0,20	0,22	0,00

Solución

Haciendo uso del lenguaje de programación R se obtuvo la configuración MDS de las poblaciones utilizando la matriz de distancias **Bhattacharyya**. Se consideró un valor $k = 6$.

Poblacion	X_1	X_2	X_3	X_4	X_5	X_6
1	0,21	0,01	−0,02	−0,06	0,07	−0,02
2	−0,02	0,13	−0,14	0,07	−0,02	0,05
3	0,17	0,06	−0,06	−0,03	0,09	−0,03
4	0,24	−0,15	0,10	−0,11	−0,14	0,10
5	−0,23	−0,12	0,12	0,06	0,13	0,14
6	−0,38	0,02	−0,01	−0,20	−0,02	−0,07
7	−0,09	−0,30	−0,05	0,14	−0,07	−0,10
8	0,06	0,18	0,26	0,07	−0,01	−0,10
9	0,15	−0,06	−0,12	−0,02	0,06	−0,03
10	−0,10	0,23	−0,08	0,07	−0,10	0,06

Solución

Se puede elegir un número apropiado de dimensiones m considerando el criterio de la proporción de varianza explicada por las primeras m dimensiones, expresada de la siguiente manera

$$\frac{\sum_{i=1}^{m}\lambda_1}{\sum_{i=1}^{n}|\lambda_i|}$$

Con ayuda del lenguaje de programación R, se calculó la varianza explicada por cada coordenada. De acuerdo a la **Figura 4**, alrededor del 80 % de la varianza es explicada cuando $k = 4$, por lo que la dimensión adecuada de la representación euclidiana es se da con un valor $k = 4$.

Por otro lado, el porcentaje de varianza explicada por las dos primeras coordenadas $k = 2$ es de aproximadamente 60 %.

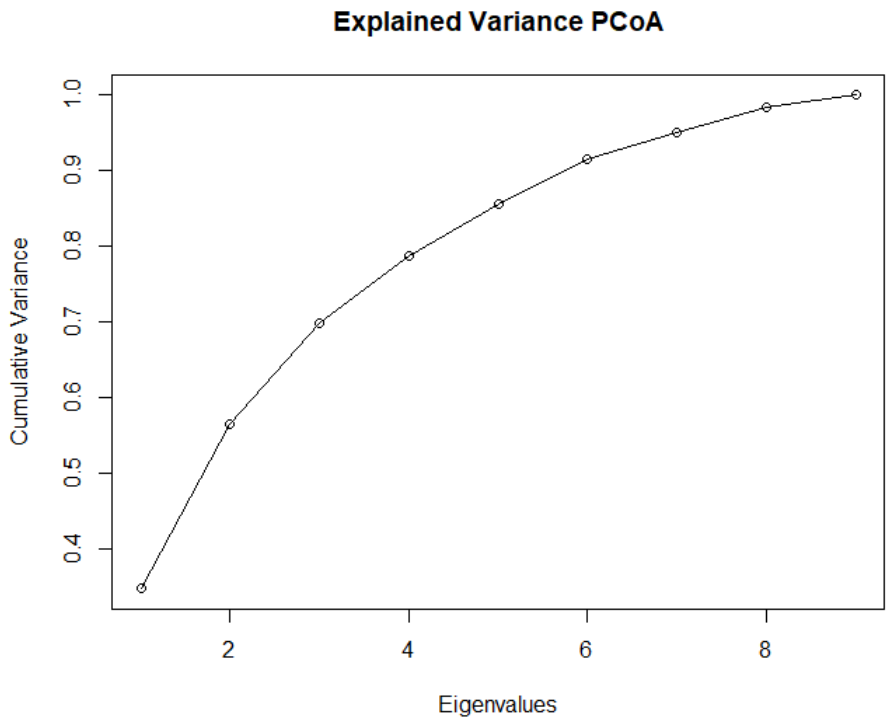

Figura 4. Varianza acumulada expliacada por cada coordenada principal.

La representación obtenida a través del modelo clásico de MDS nos muestra una clara separación entre regiones. En la esquina inferior izquierda se agrupan poblaciones de origen asiático, junto con la población eskimo. Por otro lado, en la esquina superior derecha se distinguen dos regiones claramente diferenciadas, una compuesta por poblaciones de origen africano y otra conformada por poblaciones de origen europeo (**Figura 5**).

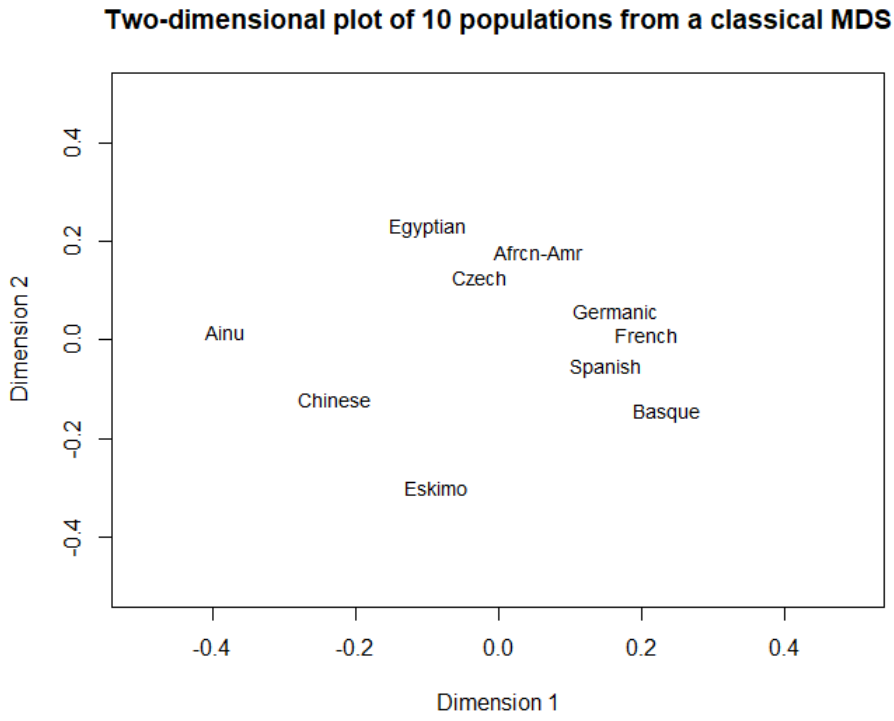

Figura 5. Configuración obtenida mediante el modelo clásico de escalamiento multidimensional.

Solución

De manera general, el patrón de agrupamiento observado en la **Figura 5** se mantiene en las representaciones obtenidas tanto por el modelo clásico como por el enfoque de mínimos cuadrados. Por otro lado, entre cada una de las transformaciones aplicadas no se observan diferencias significativas en las distancias que separan a las poblaciones (**Figura 7**).

Un criterio utilizado para la elección de la dimensionalidad adecuada es graficar el valor del 'STRESS' contra la dimensión, calculando distintas soluciones MDS para un rango de soluciones (e.g. 2 a 6 dimensiones). En la literatura se sugiere la elección de la dimensionalidad donde el valor del 'STRESS' sea menor a 0,05.

En este caso la elección de la dimensionalidad se basó en el criterio del codo de la curva, es decir, la elección de la dimensión donde la curva presente una ligera 'torcedura'. Por lo tanto, la dimensionalidad adecuada considerando la transformación de tipo *razón* sería de 4, mientras que para las transformaciones restantes *intervalo* y *ordinal*, serían de 3 y 2, respectivamente (**Figura 8**).

Solución

En la representación obtenida a través del modelo clásico MDS, la distinción entre las distintas regiones es más evidente.Por otro lado, en la representación obtenida mediante el enfoque de mínimos cuadrados, la distancia entre las poblaciones es mayor. Sin embargo, se sigue apreciando un patrón de agrupamiento coherente de acuerdo a las poblaciones consideradas (**Figura 6**).

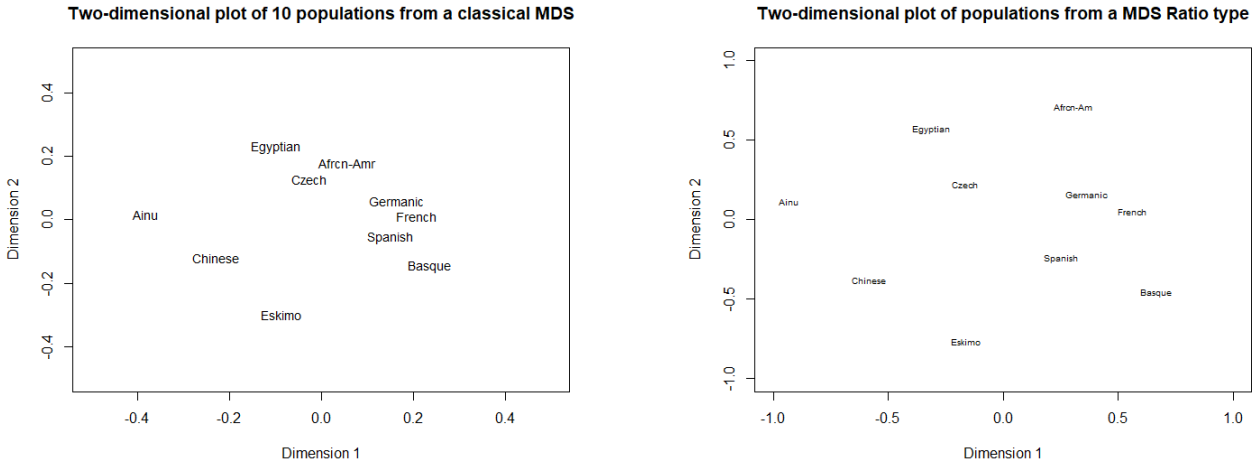

Figura 6. Comparación de la configuración obtenida mediante el modelo clásico de MDS (gráfico izquierdo) con la obtenida a través del enfoque de mínimos cuadrados (gráfico derecho).

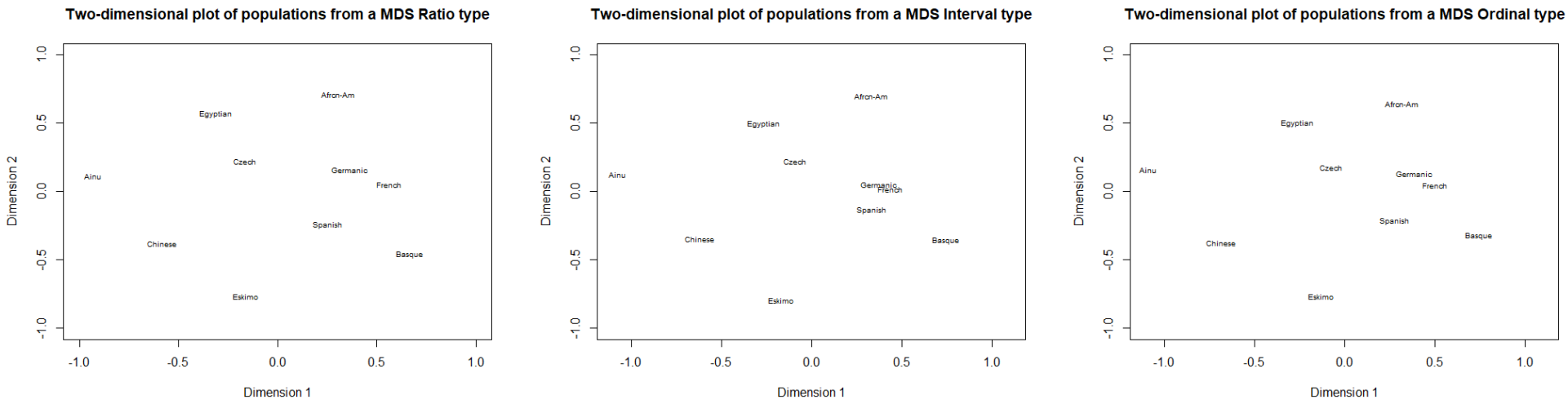

Figura 7. Configuración MDS obtenida utilizando el enfoque de mínimos cuadrados. Transformaciones de tipo razón, intervalo y ordinal para las disimilaridades.

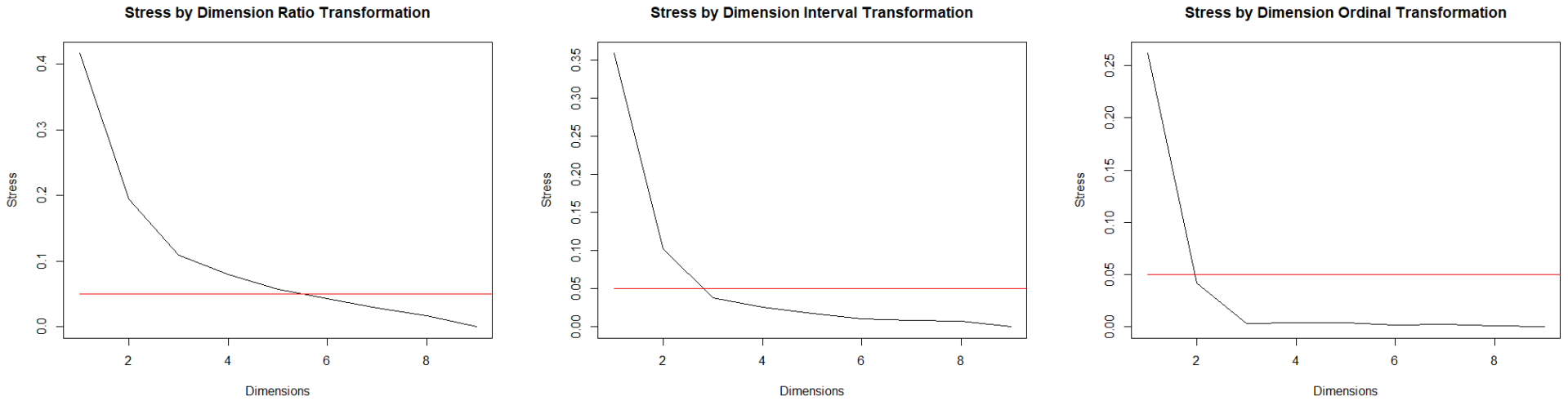

Figura 8. Stress contra dimensión. Configuración obtenida mediante el enfoque de mínimos cuadrados considerando las transformaciones de tipo razón, intervalo y ordinal para las disimilaridades.