Anales de la Real Academia de Doctores de España
Volumen 11, pp. 45-64, 2007

DISTANCIAS ENTRE LOS OPERADORES WD DE YVES ESCOUFIER EN EL ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLES*

CARLOS CASTILLA PLAZA
FCO. JAVIER DÍAZ-LLANOS SAINZ-CALLEJA
ÁNGEL FERNÁNDEZ CANCIO

OBJETIVO

El principal objetivo de este artículo es el de exponer de la forma más didáctica posible:

1. Cuatro distancias entre los operadores de las cuatro situaciones que a continuación mostramos:

$$WD\ ,\ \frac{WD}{\|WD\|},\ W^{*}D\ y\ \frac{W^{*}D}{\|W^{*}D\|}\ (1,2)$$

bajo el contexto de una **análisis de correspondencias múltiples**, en función de la **T cuadrado de Tschuprow y del número de modalidades de las parejas de las variables cualitativas en las tres primeras situaciones**, mientras que en la **cuarta** tan sólo intervendrá la **T cuadrado de Tschuprow** (aunque sabemos que para el cálculo de la **T cuadrado de Tschuprow** interviene el número de modalidades asocidos a las parejas de **variables cualitativas**). Demostraremos que la fórmula que nos permite el cálculo de las distancias entre los operadores:

$$W_1D = U_1\ D_{11}^{-1}\ U_1^{T}\ D\ y\ W_2D = U_2\ D_{12}^{-1}\ U_2^{T}D$$

asociados a los **tripletes estadísticos:**

$$\left(U_1,\ D_{11}^{-1}, D\right) y \left(U_2,\ D_{12}^{-1}, D\right),\ respectivamente.$$

es la misma que la distancia entre los operadores:

$$W_1^{*}D = U_1^{*}\ D_{11}^{-1}\ U_1^{*T}\ D\ y\ W_2^{*}D = U_2^{*}\ D_{12}^{-1}\ U_2^{*T}\ D$$

* Conferencia pronunciada el 19 de junio de 2007 en la Real Academia de Doctores de España.

asociados a los **tripletes estadísticos:**

$$\left(U_1^*,\ D_{I1}^{-1}, D \right) y \left(U_2^*,\ D_{I2}^{-1}, D \right), \textit{ respectivamente.}$$

2. Un ejercicio didáctico y otro práctico extraidos de (1,2) y (4) con el fin de manejar los conceptos ya definidos en (1,2).

Dado que las matrices de datos no son de grandes dimensiones podremos llegar, sin dificultad, a las **matrices de distancias entre operadores** mediante una simple calculadora.

Palablas clave: Operadores WD de Yves Escoufier bajo un contexto de un análisis de correspondencias múltiple, distancia entre los operadores WD, T cuadrado de Tschuprow.

INTRODUCCIÓN

El hecho de que no se encuentre el **cálculo de las distancias entre operadores** en ningún libro de Análisis Estadístico Multidimensional escrito en español nos ha motivado contemplarlo en este artículo por su gran utilidad práctica.

A partir de los **tripletes estadísticos** definidos por la **matriz de variables indicadoras** asociadas a cada una de las **variables cualitativas,** la **métrica que nos permite** el cálculo de las distancias entre los individuos y la **métrica que nos permite** el cálculo de las distancias entre las modalidades de cada **variable cualitativa** podemos construir los operadores y, a partir de ellos sus distancias.

MATERIAL Y MÉTODO

Basándonos en los conceptos contenidos en (1,2) vamos a proceder:

1. Al cálculo de las distancias entre los operadores.

2. A la aplicación de las fórmulas obtenidas a un ejemplo didáctico y otro práctico extraidos de (1,2) y (4), respectivamente.

Mientras que en el ejercicio didáctico calcularemos las cuatro matrices de distancias entre operadores obtenidas mediante las fórmulas del apartado 1, en el ejercicio práctico mostraremos una estrategia metodológica y además calcularemos la matriz de distancias entre los operadores transformados y normados.

Mientras que los resultados del ejercicio didáctico carecen de interpretación por las dimensiones de la matriz, los del ejercicio práctico si se podran interpretar dado que la matriz es de dimensiones(19 x 15). En este sentido, Thiery Foucart (3) comenta que un **análisis de componentes principales** dará resultados interesantes si el número de individuos es superior a 15 y el de variables a 4.

CÁLCULO DE LAS DISTANCIAS ENTRE OPERADORES

1. Distancia entre los operadores:

$$W_1 D \text{ y } W_2 D$$

Para la pareja de operadores:

$$W_1 D \text{ y } W_2 D$$

la distancia entre ellos se define de la siguiente manera:

$$d(W_1 D, W_2 D) = \sqrt{Tr(W_1 D - W_2 D)^2} = \sqrt{Tr(W_1 D)^2 + Tr(W_2 D)^2 - 2Tr(W_1 D W_2 D)}$$

Sustituyendo los resultados obtenidos –referentes a las trazas de los operadores en (1,2)– en la fórmula de la distancia entre los operadores, tenemos que:

$$d(W_1 D, W_2 D) = \sqrt{m_1 + m_2 - 2\left[\sum_{i1=1}^{i1=m_1} \sum_{i2=1}^{i2=m_2} \frac{\left[p(i1,i2) - p_{i1+}\ p_{+i2}\right]^2}{p_{i1+}\ p_{+i2}} + 1\right]}$$

La traducción de esta relación en términos de estimaciones espíricas es la que a continuación mostramos:

$$d(W_1 D, W_2 D) = \sqrt{m_1 + m_2 - 2\left[\sum_{i1=1}^{i1=m_1} \sum_{i2=1}^{i2=m_2} \frac{\left[\frac{n(i1,i2)}{n} - \frac{n_{i1+}}{n}\ \frac{n_{+i2}}{n}\right]^2}{\frac{n_{i1+}}{n}\ \frac{n_{+i2}}{n}} + 1\right]}$$

Operando convenientemente llegamos a una expresión más operativa:

$$d(W_1 D, W_2 D) = \sqrt{m_1 + m_2 - 2\left[\sum_{i1=1}^{i1=m_1} \sum_{i2=1}^{i2=m_2} \frac{\left[n(i1,i2)\right]^2}{n_{i1+}\ n_{+i2}} - 1 + 1\right]}$$

Dado que:

$$\sum_{i1=1}^{i1=m_1} \sum_{i2=1}^{i2=m_2} \frac{\left[n(i1,i2)\right]^2}{n_{i1+}\ n_{+i2}} - 1 = \sqrt{(m_1 - 1)(m_2 - 1)}\ T^2_{I1,I2}$$

tenemos,

$$d(W_1 D, W_2 D) = \sqrt{m_1 + m_2 - 2\left[\sqrt{(m_1-1)(m_2-1)}\; T^2_{I1,I2} + 1\right]}$$

Por tanto,

$$d(W_1 D, W_2 D) = \sqrt{(m_1-1)+(m_2-1) - 2\sqrt{(m_1-1)(m_2-1)}\; T^2_{I1,I2}}$$

Operando de la misma manera obtendremos, sin dificultad, las otras dos distancias entre operadores:

$$d\left(W_1 D, W_3 D\right) = \sqrt{(m_1-1)+(m_3-1) - 2\sqrt{(m_1-1)(m_3-1)}\; T^2_{I1,I3}}$$

$$d\left(W_2 D, W_3 D\right) = \sqrt{(m_2-1)+(m_3-1) - 2\sqrt{(m_2-1)(m_3-1)}\; T^2_{I2,I3}}$$

2. Distancia entre los operadores:

$$\frac{W_1 D}{\|W_1 D\|} \; y \; \frac{W_2 D}{\|W_2 D\|}$$

Para la pareja de operadores:

$$\frac{W_1 D}{\|W_1 D\|} \; y \; \frac{W_2 D}{\|W_2 D\|}$$

la distancia entre ellos se define de la siguiente forma:

$$d\left(\frac{W_1 D}{\|W_1 D\|}, \frac{W_2 D}{\|W_2 D\|}\right) = \sqrt{Tr\left(\frac{W_1 D}{\|W_1 D\|} - \frac{W_2 D}{\|W_2 D\|}\right)^2} =$$

$$= \sqrt{Tr\left(\frac{W_1 D}{\|W_1 D\|}\right)^2 + Tr\left(\frac{W_2 D}{\|W_2 D\|}\right)^2 - 2\,Tr\left(\frac{W_1 D}{\|W_1 D\|}\;\frac{W_2 D}{\|W_2 D\|}\right)}$$

Teniendo en cuenta el siguiente resultado

$$\|W_1 D\| = \sqrt{Tr(W_1 D)} \quad \|W_2 D\| = \sqrt{Tr(W_2 D)}$$

y, haciendo uso de la definición del **coeficiente RV** (1,2), tenemos que:

$$d\left(\frac{W_1D}{\|W_1D\|},\frac{W_2D}{\|W_2D\|}\right)= \sqrt{1+1-2\,RV(W_1D,W_2D)} = \sqrt{2\left[\,1\,-\,RV(W_1D,W_2D)\right]}$$

Sustituyendo los resultados obtenidos en (1,2) obtenemos la siguiente expresión:

$$d\left(\frac{W_1D}{\|W_1D\|},\frac{W_2D}{\|W_2D\|}\right)= \sqrt{2\left(1\,-\left(\frac{\displaystyle\sum_{i1=1}^{i1=m_1}\sum_{i2=1}^{i2=m_2}\frac{[p(i1,i2)\,-\,p_{i1+}p_{+i2}]}{p_{i1+}\;p_{+i2}}+1}{\sqrt{m_1\;m_2}}\right)\right)}$$

Operando convenientemente deducimos la expresión de la segunda distancia:

$$d\left(\frac{W_1D}{\|W_1D\|},\frac{W_2D}{\|W_2D\|}\right)= \sqrt{2\left[1\,-\left(\frac{\sqrt{(m_1-1)(m_2-1)}\;\;T^2_{I1,I2}\;+\;1}{\sqrt{m_1\;m_2}}\right)\right]}$$

Operando de la misma manera obtendremos, sin dificulad, las otras dos distancias entre operadores:

$$d\left(\frac{W_1D}{\|W_1D\|},\frac{W_3D}{\|W_3D\|}\right)= \sqrt{2\left[1\,-\left(\frac{\sqrt{(m_1-1)(m_3-1)}\;\;T^2_{I1,I3}+\;1}{\sqrt{m_1\;m_3}}\right)\right]}$$

$$d\left(\frac{W_2D}{\|W_2D\|},\frac{W_3D}{\|W_3D\|}\right)= \sqrt{2\left[1\,-\left(\frac{\sqrt{m_2-1)(m_3-1)}\;\;T^2_{I2,I3}+\;1}{\sqrt{m_2\;m_3}}\right)\right]}$$

3. Distancia entre los operadores:

$$W^*_1D \text{ y } W^*_2D$$

Para la pareja de operadores:

$$W^*_1D \text{ y } W^*_2D$$

la distancia entre ellos se define de la siguiente forma:

$$d\left(W_1^* D, W_2^* D\right) = \sqrt{Tr\left(W_1^* D \; - \; W_2^* D\right)^2} =$$

$$= \sqrt{Tr\left(W_1^* D\right)^2 + Tr\left(W_2^* D\right)^2 - 2\, Tr\left(W_1^* D \; W_2^* D\right)}$$

Sustituyendo los resultados obtenidos –referentes a las trazas de los operadores en (1,2)– en la fórmula de la distancia entre los operadores transformados, tenemos que:

$$d\left(W_1^* D, W_2^* D\right) = \sqrt{(m_1 - 1) + (m_2 - 1) - 2 \sum_{i1=1}^{i1=m_1} \sum_{i2=1}^{i2=m_2} \frac{\left[p(i1,i2) - p_{i1+}\; p_{+i2}\right]^2}{p_{i1+}\; p_{+i2}}}$$

La traducción de esta relación en términos de estimaciónes es la que a continuación mostramos:

$$d\left(W_1^* D, W_2^* D\right) = \sqrt{(m_1 - 1) + (m_2 - 1) - 2 \sum_{i1=1}^{i1=m_1} \sum_{i2=1}^{i2=m_2} \frac{\left[\frac{n(i1,i2)}{n} - \frac{n_{i1+}}{n}\; \frac{n_{+i2}}{n}\right]^2}{\frac{n_{i1+}}{n}\; \frac{n_{+i2}}{n}}}$$

Operando convenientemente llegamos a una expresión más operativa:

$$d\left(W_1^* D, W_2^* D\right) = \sqrt{(m_1 - 1) + (m_2 - 1) - 2 \left[\sum_{i1=1}^{i1=m_1} \sum_{i2=1}^{i2=m_2} \frac{\left[n(i1,i2)\right]^2}{n_{i1+}\; n_{+i2}} - 1\right]}$$

Dado que:

$$\sum_{i1=1}^{i1=m_1} \sum_{i2=1}^{i2=m_2} \frac{\left[n(i1,i2)\right]^2}{n_{i1+}\; n_{+i2}} - 1 = \sqrt{(m_1 - 1)(m_2 - 1)}\;\; T^2_{I1,I2}$$

tenemos,

$$d\left(W_1^* D, W_2^* D\right) = \sqrt{(m_1 - 1) + (m_2 - 1) - 2\; \sqrt{(m_1 - 1)(m_2 - 1)}\;\; T^2_{I1,I2}}$$

Operando de la misma manera obtendremos, sin dificultad, las otras dos distancias entre operadores:

$$d\left(W_1^* D, W_3^* D\right) = \sqrt{(m_1 - 1) + (m_3 - 1) - 2\; \sqrt{(m_1 - 1)(m_3 - 1)}\;\; T^2_{I1,I3}}$$

$$d\left(W_2^* D, W_3^* D\right) = \sqrt{(m_2 - 1)(m_3 - 1) - 2\; \sqrt{(m_2 - 1)(m_3 - 1)}\;\; T^2_{I2,I3}}$$

4. Distancia entre los operadores:

$$\frac{W_1^* D}{\left\| W_1^* D \right\|} \; y \; \frac{W_2^* D}{\left\| W_2^* D \right\|}$$

Para la pareja de operadores:

$$\frac{W_1^* D}{\left\| W_1^* D \right\|} \; y \; \frac{W_2^* D}{\left\| W_2^* D \right\|}$$

la distancia entre ellos se define de la siguiente forma:

$$d\left(\frac{W_1^* D}{\left\| W_1^* D \right\|}, \frac{W_2^* D}{\left\| W_2^* D \right\|} \right) = \sqrt{Tr\left(\frac{W_1^* D}{\left\| W_1^* D \right\|} - \frac{W_2^* D}{\left\| W_2^* D \right\|} \right)^2} =$$

$$= \sqrt{Tr\left(\frac{W_1^* D}{\left\| W_1^* D \right\|} \right)^2 + Tr\left(\frac{W_2^* D}{\left\| W_2^* D \right\|} \right)^2 - 2\, Tr\left(\frac{W_1^* D}{\left\| W_1^* D \right\|} \; \frac{W_2^* D}{\left\| W_2^* D \right\|} \right)}$$

Sustituyendo los resultados obtenidos –referentes a las trazas de los operadores en (1,2)– en la fórmula de la distancia entre los operadores WD transformados y normados, tenemos que:

$$d\left(\frac{W_1^* D}{\left\| W_1^* D \right\|}, \frac{W_2^* D}{\left\| W_2^* D \right\|} \right) = \sqrt{2\left(1 - T_{I1,I2}^2\right)}$$

Operando de la misma manera obtendremos, sin dificultad, las otras dos distancias entre operadores:

$$d\left(\frac{W_1^* D}{\left\| W_1^* D \right\|}, \frac{W_3^* D}{\left\| W_3^* D \right\|} \right) = \sqrt{2\left(1 - T_{I1,I3}^2\right)}$$

$$d\left(\frac{W_2^* D}{\left\| W_2^* D \right\|}, \frac{W_3^* D}{\left\| W_3^* D \right\|} \right) = \sqrt{2\left(1 - T_{I2,I3}^2\right)}$$

Observaciones:

1. Así como la distancia entre los operadores

$$WD \; y \; W^* D$$

es la misma, la distancia entre los operadores

$$\frac{WD}{\left\| WD \right\|} \; y \; \frac{W^* D}{\left\| W^* D \right\|}$$

no es la misma.

2. Entre las **cuatro distancias** que hemos deducido, la que utilizaremos cuando deseemos contruir la matriz de distancias entre operadores con el fin de obtener un **dendrograma** será la distancia entre operadores transformados y normados.

3. Así como se demuestra facílmente que la distancia entre dos variables:

$$y_j = \frac{x_{ij} - \overline{x}_j}{s_j \sqrt{n}} \quad e \quad y_j{}' = \frac{x_{ij'} - \overline{x}_{j'}}{s_{j'} \sqrt{n}}$$

donde

$$\overline{x}_j \;\; y \;\; \overline{x}_{j'}$$

representan las medias de dos variables de la **tabla de datos cuantitativos** y

$$s_j \;\; y \;\; s_{j'}$$

representan las desviaciones típicas, respectivamente, es igual a:

$$d\left(y_j, y_{j'}\right) = \sqrt{2\left(1 - r_{x_j, x_{j'}}\right)}$$

donde:

$$r_{j,j'}$$

es el **coeficiente de correlación lineal de Bravais-Pearson** entre variables cuantitativas contenidas en la **tabla de datos cuantitativos**, la distancia entre operadores transformados y normados presenta una pequeña variación en cuanto a ésta y es que en lugar del **coeficiente de correlación lineal de Bravais-Pearson** entre dos variables cuantitativas contiene el **coeficiente T cuadrado de Tschuprow.**

Aplicación de las fórmulas obtenidas a un ejercicio didáctico y a un ejercicio práctico

1. Ejercicio didáctico para el cálculo de las matrices de las distancias entre los operadores de Yves Escoufier.

Para la construcción de las matrices de distancias entre operadores hemos tenido en cuenta las tres **T cuadrado de Tschuprow** obtenidas en el ejercicio didáctico (1,2).

$$T^2_{I1,I2} = 0,2357 \quad T^2_{I1,I3} = 0,5774 \quad T^2_{I2,I3} = 0,8165$$

$$T^2_{I1,I1} = T^2_{I2,I2} = T^2_{I3,I3} = 1$$

Primera situación:

Cálculo de la matriz de distancias entre los operadores WD

Aplicando la fórmula ya obtenida a nuestros datos provenientes del ejercicio didáctico (1,2) llegamos- sin dificultad- a los resultados que nos permiten construir la matriz de distancias entre los operadores de la siguiente manera:

$$d\left(W_1 D, W_2 D\right) = \sqrt{(1)+(2)-2\sqrt{(1)(2)}\,0,2357} = 1,5275$$

$$d\left(W_1 D, W_2 D\right) = 1,5275$$

$$d\left(W_1 D, W_3 D\right) = \sqrt{(1)+(3)-2\sqrt{(1)(3)}\,0,5774} = 1,4142$$

$$d\left(W_1 D, W_3 D\right) = 1,4142$$

$$d\left(W_2 D, W_3 D\right) = \sqrt{(2)+(3)-2\sqrt{(2)(3)}\,0,8165} = 1,0000$$

$$d\left(W_2 D, W_3 D\right) = 1,0000$$

$$d\left(W_1 D, W_1 D\right) = d\left(W_2 D, W_2 D\right) = d\left(W_3 D, W_3 D\right) = 0$$

Por consiguiente, la matriz de distancias entre los operadores WD es:

$$\begin{pmatrix} 0,0000 & 1,5275 & 1,4142 \\ 1,5275 & 0,0000 & 1,0000 \\ 1,4142 & 1,0000 & 0,0000 \end{pmatrix}$$

Segunda situación:

Cálculo de la matriz de distancias entre los operadores WD normados

Aplicando la fórmula ya obtenida a nuestros datos provenientes del ejercicio didáctico (1,2) llegamos- sin dificultad- a los resultados que nos permiten construir la matriz de distancia entre los operadores de la siguiente manera:

$$d\left(\frac{W_1D}{\|W_1D\|},\frac{W_2D}{\|W_2D\|}\right)=\sqrt{2\left[1-\left(\frac{\sqrt{(1)(2)}\;0,2357\;+\;1}{\sqrt{(2)((3)}}\right)\right]}=\;0,9546$$

$$d\left(\frac{W_1D}{\|W_1D\|},\frac{W_2D}{\|W_2D\|}\right)=\;0,9546$$

$$d\left(\frac{W_1D}{\|W_1D\|},\frac{W_3D}{\|W_3D\|}\right)=\sqrt{2\left[1-\left(\frac{\sqrt{(1)(3)}\;0,5774+1}{\sqrt{(2)(4)}}\right)\right]}=\;0,7653$$

$$d\left(\frac{W_1D}{\|W_1D\|},\frac{W_3D}{\|W_3D\|}\right)=\;0,7653$$

$$d\left(\frac{W_2D}{\|W_2D\|},\frac{W_3D}{\|W_3D\|}\right)=\sqrt{2\left[1\;-\;\left(\frac{\sqrt{(2)(3)}\;0,8165+1}{(3)(4)}\right)\right]}=0,5176$$

$$d\left(\frac{W_2D}{\|W_2D\|},\frac{W_3D}{\|W_3D\|}\right)=\;0,5176$$

$$d\left(\frac{W_1D}{\|W_1D\|},\frac{W_1D}{\|W_1D\|}\right)=d\left(\frac{W_2D}{\|W_2D\|},\frac{W_2D}{\|W_2D\|}\right)=d\left(\frac{W_3D}{\|W_3D\|},\frac{W_3D}{\|W_3D\|}\right)=0$$

Por consiguiente, la matriz de distancias entre los operadores WD normados es:

$$\begin{pmatrix}0,0000 & 0,9546 & 0,7653\\ 0,9546 & 0,0000 & 0,5176\\ 0,7653 & 0,5176 & 0,0000\end{pmatrix}$$

Tercera situación:

Cálculo de la matriz de distancias entre los operadores WD transformados

Dado que la fórmula que se obtiene para el cálculo de la distancia entre operadores es la misma que en la **primera situación** no es necesario exponer de nuevo los mismos resultados.

Cuarta situación:

Cálculo de la matriz de distancias entre los operadores WD transformados y normados

Aplicando la fórmula ya obtenida en a nuestros datos provenientes del ejercicio didáctico (1,2)- llegamos sin dificultad- a los resultados que nos permiten construir la matriz de distancias entre los operadores de la siguiente manera:

$$d\left(\frac{W_1^*D}{\|W_1^*D\|},\frac{W_2^*D}{\|W_2^*D\|}\right)=\sqrt{2\left(1-0,2357\right)}=1,2364$$

$$d\left(\frac{W_1^*D}{\|W_1^*D\|},\frac{W_2^*D}{\|W_2^*D\|}\right)=1,2364$$

$$d\left(\frac{W_1^*D}{\|W_1^*D\|},\frac{W_3^*D}{\|W_3^*D\|}\right)=\sqrt{2(1-0,5774)}=0,9193$$

$$d\left(\frac{W_1^*D}{\|W_1^*D\|},\frac{W_3^*D}{\|W_3^*D\|}\right)=0,9193$$

$$d\left(\frac{W_2^*D}{\|W_2^*D\|},\frac{W_3^*D}{\|W_3^*D\|}\right)=\sqrt{2(1-0,8165)}=0,6068$$

$$d\left(\frac{W_2^*D}{\|W_2^*D\|},\frac{W_3^*D}{\|W_3^*D\|}\right)=0,6058$$

$$d\left(\frac{W_1^*D}{\|W_1^*D\|},\frac{W_1^*D}{\|W_1^*D\|}\right)=d\left(\frac{W_2^*D}{\|W_2^*D\|},\frac{W_2^*D}{\|W_2^*D\|}\right)=d\left(\frac{W_3^*D}{\|W_3^*D\|},\frac{W_3^*D}{\|W_3^*D\|}\right)=0$$

Por consiguiente, la matriz de distancias entre los operadores transformados y normados es:

$$\begin{pmatrix} 0,0000 & 1,2364 & 0,9193 \\ 1,2364 & 0,0000 & 0,6058 \\ 0,9193 & 0,6058 & 0,0000 \end{pmatrix}$$

2. Ejercicio práctico para el cálculo de las distancias entre operadores transformados y normados.

La **tabla de datos cuantitativos** está constituida por 19 panes vendidos en un envoltorio de plástico y por cinco variables cuantitativas: precio/100gr (PRI), contenido en agua (EAU), contenido en glúcidos (GLU), contenido en proteinas (PRO) y cantidad de calorias/100gr(CAL).

La tabla de datos ha sido extraida del libro de E. Diday, J. Lemaire, J. Pouget y F. Testu (4).

La estrategia metodológica que hemos llevado a cabo para llegar al resultado final;es decir, a la matriz de distancias entre los operadores transformados y normados es la que mostramos a continuación:

Primera etapa: transformación de una tabla de datos cuantitativos heterogenea a una tabla disyuntiva completa.

Segunda etapa: construcción de k(k-1)/2 tablas de contingencias.

Tercera etapa: construcción de k(k-1)/2 T cuadrado de Tschuprow.

Cuarta etapa: construcción de k(k-1)/2 distancias entre operadores transformados y normados

Primera etapa: transformación de una tabla de datos cuantitativos heterogenea en una tabla disyuntiva completa.

Los pasos a seguir son los siguientes:

1.º Presentación de la **tabla de datos cuantitativos** y un breve comentario sobre ella.

2.º Construcción de la **tabla numérica**.

3.º Construcción de la **tabla disyuntiva completa**.

1.º Presentación de la **tabla de datos cuantitativos** y un breve comentario sobre ella.

TABLA DE DATOS CUANTITATIVOS

	PRI	*EAU*	*GLU*	*PRO*	*CAL*
1	63	3635	4940	875	266
2	77	3190	5285	855	287
3	86	3515	5025	845	434
4	86	3530	5220	820	262
5	89	3350	4850	1150	280
6	91	3070	5260	905	293
7	92	3130	5280	920	290
8	92	3635	5030	810	266
9	95	3490	5010	900	276
10	95	3460	5135	830	275
11	106	3380	5230	930	268
12	74	3500	5330	860	259
13	76	4030	4880	755	238
14	85	3365	5415	830	262
15	57	3515	4970	1000	264
16	95	3960	4550	895	253
17	132	2925	5585	945	292
18	152	2720	5230	1030	326
19	153	2340	5410	930	358

Comentario: de la mera observación de la **tabla de datos cuantitativos** concluimos que es **heterogenea.**

2.º Construcción de la **tabla numérica.**

Debido al **carácter heterogéneo** de la **tabla de datos cuántitativos** vamos a reemplazar las **variables cuántitativas** iniciales por **variables cualitativas a tres modalidades** tal como se indica en (4).

TABLA NUMÉRICA

PRI	*EAU*	*GLU*	*PRO*	*CAL*
1	3	1	2	1
1	1	3	2	3
1	3	2	1	3
1	3	2	1	1
2	2	1	3	2
2	1	2	2	3
2	1	3	3	3
2	3	2	1	1
3	2	1	2	2
3	2	2	1	2
3	2	2	3	2
1	3	3	2	1
1	3	1	1	1
1	2	3	1	1
1	3	1	3	1
3	3	1	2	1
3	1	3	3	3
3	1	2	3	3
3	1	3	3	3

3.º Construcción de la **tabla disyuntiva completa**

TABLA DISYUNTIVA COMPLETA

PRI			*EAU*			*GLU*			*PRO*			*CAL*		
PRI1	*PRI2*	*PRI3*	*EAU1*	*EAU2*	*EAU3*	*GLU1*	*GLU2*	*GLU3*	*PRO1*	*PRO2*	*PRO3*	*CAL1*	*CAL2*	*CAL3*
1	0	0	0	0	1	1	0	0	0	1	0	1	0	0
1	0	0	1	0	0	0	0	1	0	1	0	0	0	1
1	0	0	0	0	1	0	1	0	1	0	0	0	0	1
1	0	0	0	0	1	0	1	0	1	0	0	1	0	0
0	1	0	0	1	0	1	0	0	0	0	1	0	1	0
0	1	0	1	0	0	0	1	0	0	1	0	0	0	1
0	1	0	1	0	0	0	0	1	0	0	1	0	0	1
0	1	0	0	0	1	0	1	0	1	0	0	1	0	0
0	0	1	0	1	0	1	0	0	0	1	0	0	1	0
0	0	1	0	1	0	0	1	0	1	0	0	0	1	0
0	0	1	0	1	0	0	1	0	0	0	1	0	1	0
1	0	0	0	0	1	0	0	1	0	1	0	1	0	0
1	0	0	0	0	1	1	0	0	1	0	0	1	0	0
1	0	0	0	1	0	0	0	1	1	0	0	1	0	0
1	0	0	0	0	1	1	0	0	0	0	1	1	0	0
0	0	1	0	0	1	1	0	0	0	1	0	1	0	0
0	0	1	1	0	0	0	0	1	0	0	1	0	0	1
0	0	1	1	0	0	0	1	0	0	0	1	0	0	1
0	0	1	1	0	0	0	0	1	0	0	1	0	0	1

De la mera observación de la **tabla de disyuntiva completa** concluimos que no es necesario la eliminación de ninguna columna ya que el número de unos en cada **variable indicadora** es superior al 10%.

Segunda etapa: construcción de k(k-1)/2 tablas de contingencia

En nuestro caso concreto tendremos que construir 10 **tablas de contingencias.**

	EAU1	*EAU2*	*EAU3*	
PRI1	1	1	6	8
PRI2	2	1	1	4
PRI3	3	3	1	7
	6	5	8	19

	GLU1	*GLU2*	*GLU3*	
PRI1	3	2	3	8
PRI2	1	2	1	4
PRI3	2	3	2	7
	6	7	6	19

	PRO1	*PRO2*	*PRO3*	
PRI1	4	3	1	8
PRI2	1	1	2	4
PRI3	1	2	4	7
	6	6	7	19

	CAL1	*CAL2*	*CAL3*	
PRI1	6	0	2	8
PRI2	1	1	2	4
PRI3	1	3	3	7
	8	4	7	19

	GLU1	*GLU2*	*GLU3*	
EAU1	0	2	4	6
EAU2	2	2	1	5
EAU3	4	3	1	8
	6	7	6	19

	PRO1	*PRO2*	*PRO3*	
EAU1	0	2	4	6
EAU2	2	1	2	5
EAU3	4	3	1	8
	6	6	7	19

	CAL1	*CAL2*	*CAL3*	
EAU1	0	0	6	6
EAU2	1	4	0	5
EAU3	7	0	1	8
	8	4	7	19

	PRO1	*PRO2*	*PRO3*	
GLU1	1	3	2	6
GLU2	4	1	2	7
GLU3	1	2	3	6
	6	6	7	19

	CAL1	*CAL2*	*CAL3*	
GLU1	4	2	0	6
GLU2	2	2	3	7
GLU3	2	0	4	6
	8	4	7	19

	CAL1	*CAL2*	*CAL3*	
PRO1	4	1	1	6
PRO2	3	1	2	6
PRO3	1	2	4	7
	8	4	7	19

Tercera etapa: cálculo de las k(k-1)/2 T cuadrado de Tschuprow.

A partir de las **tablas de contingencias** podemos calcular facilmente las **T cuadrado de Tschuprow** aplicando la fórmula ya contenida en (1,2).

Así como la matriz de distancias contendrá cuatro decimales la matriz **T cuadrado de Tschuprow** contendrá 8 ya que pequeñas variaciones en ésta matriz conlleva grandes variaciones en el cálculo de los valores propios y por consiguiente, el redondeo no es adecuado.

A continuación mostramos los valores de las 10 **T cuadrado de Tschuprow.**

$$T^2_{PRI,EAU} = 0,17276786 \qquad T^2_{PRI,GLU} = 0,02338435$$

$$T^2_{PRI,PRO} = 0,10522959 \qquad T^2_{PRI,CAL} = 0,19674745$$

$$T^2_{EAU,GLU} = 0,16775794 \qquad T^2_{EAU,PRO} = 0,15585317$$

$$T^2_{EAU,CAL} = 0,73281250 \qquad T^2_{GLU,PRO} = 0,10629252$$

$$T^2_{GLU,CAL} = 0,18112245 \qquad T^2_{PRO,CAL} = 0,10522959$$

Interpretación de la T cuadrado de Tschuprow

En el capítulo XIV del libro de Caillez,F y Pages,J-P (5) se hace una referencia en cuanto a la interpreción de la **T cuadrado de Tschuprow**. Dicha referencia ha sido extraida del libro de Kendall,M.G y Stuart,A(6). Estos autores dicen que el coeficiente **T cuadrado de Tschuprow** sirve para medir el grado de asociación entre dos variables cualitativas. Dicho grado de asociación varía de 0 a 1, siendo mayor a medida que se aproxima a 1.

Por consiguiente,

1. Las variables cualitativas que presentan mayor grado de asociación son el contenido en agua (EAU) con la cantidad de calorias (CAL) o viceversa.

2. Las variables cualitativas que presentan menor grado de asociación son el precio (PRI) con el contenido el glúcidos (GLU) o viceversa.

3. Existen dos parejas de variables que presentan un grado de asociación algo mayor que el precio (PRI) con el contenido en glúcidos (GLU).

Estas parejas de variables son: el precio (PRI) con el contenido en proteinas (PRO) y el contenido en proteinas (PRO) con la cantidad de calorías (CAL).

Cuarta etapa: construcción de k(k-1)/2 distancias entre operadores transformados y normados

Aplicando la fórmula que relaciona la distancia entre los operadores transformados y normados con la **T cuadrado de Tschuprow** a nuestro caso concreto obtenemos los siguientes resultados:

$$d\left(\frac{W_1^* D}{\|W_1^* D\|}, \frac{W_2^* D}{\|W_2^* D\|}\right) = \sqrt{2\left(1 - T^2_{PRI,EAU}\right)} = 1{,}2863$$

$$d\left(\frac{W_1^* D}{\|W_1^* D\|}, \frac{W_3^* D}{\|W_3^* D\|}\right) = \sqrt{2\left(1 - T^2_{PRI,GLU}\right)} = 1{,}3976$$

$$d\left(\frac{W_1^* D}{\|W_1^* D\|}, \frac{W_4^* D}{\|W_4^* D\|}\right) = \sqrt{2\left(1 - T^2_{PRI,PRO}\right)} = 1{,}3377$$

$$d\left(\frac{W_1^* D}{\|W_1^* D\|}, \frac{W_5^* D}{\|W_5^* D\|}\right) = \sqrt{2\left(1 - T^2_{PRI,CAL}\right)} = 1{,}2675$$

$$d\left(\frac{W_2^* D}{\|W_2^* D\|}, \frac{W_3^* D}{\|W_3^* D\|}\right) = \sqrt{2\left(1 - T^2_{EAU,GLU}\right)} = 1{,}2901$$

$$d\left(\frac{W_2^* D}{\|W_2^* D\|}, \frac{W_4^* D}{\|W_4^* D\|}\right) = \sqrt{2\left(1 - T^2_{EAU,PRO}\right)} = 1{,}2993$$

$$d\left(\frac{W_2^* D}{\|W_2^* D\|}, \frac{W_5^* D}{\|W_5^* D\|}\right) = \sqrt{2\left(1 - T^2_{EAU,CAL}\right)} = 0{,}7310$$

$$d\left(\frac{W_3^* D}{\|W_3^* D\|}, \frac{W_4^* D}{\|W_4^* D\|}\right) = \sqrt{2\left(1 - T^2_{GLU,PRO}\right)} = 1{,}3369$$

$$d\left(\frac{W_3^* D}{\|W_3^* D\|}, \frac{W_5^* D}{\|W_5^* D\|}\right) = \sqrt{2\left(1 - T^2_{GLU,CAL}\right)} = 1{,}2797$$

$$d\left(\frac{W_4^* D}{\|W_4^* D\|}, \frac{W_5^* D}{\|W_5^* D\|}\right) = \sqrt{2\left(1 - T^2_{PRO,CAL}\right)} = 1{,}3377$$

Por consiguiente, la matriz de distancias entre operadores transformados y normados es:

$$
\begin{array}{c}
\\ 1 \\ 2 \\ 3 \\ 4 \\ 5
\end{array}
\begin{array}{c}
\begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{pmatrix}
0,0000 & 1,2863 & 1,3976 & 1,3377 & 1,2675 \\
1,2863 & 0,0000 & 1,2901 & 1,2993 & 0,7310 \\
1,3976 & 1,2901 & 0,0000 & 1,3369 & 1,2797 \\
1,3377 & 1,2993 & 1,3369 & 0,0000 & 1,3377 \\
1,2675 & 0,7310 & 1,2797 & 1,3377 & 0,0000
\end{pmatrix}
\end{array}
$$

$$1: PRI - \frac{W_1^* D}{\|W_1^* D\|} \quad 2: EAU - \frac{W_2^* D}{\|W_2^* D\|} \quad 3: GLU - \frac{W_3^* D}{\|W_3^* D\|}$$

$$4: PRO - \frac{W_4^* D}{\|W_4^* D\|} \quad 5: CAL - \frac{W_5^* D}{\|W_5^* D\|}$$

CONCLUSIONES

– Cuando partamos de una matriz de datos cuantitativos heterogenea es aconsejable pasar las variables cuantitativas iniciales a cualitativas. Es deseable que el número de modalidades no sea elevado y a su vez sea el mismo.

– La distancia entre los operadores:

$$WD \text{ y } W^* D$$

es la misma.

– Cuando deseemos construir la matriz de distancia entre operadores la opción más indicada es la distancia entre los operadores transformados y normados:

$$\frac{W^* D}{\|W^* D\|}$$

BIBLIOGRAFÍA

(1) Díaz-Llanos Sainz-Calleja, Fco. J. (1995). El tratamiento estadístico de las encuestas de opinión, pieza clave en la ingeniería de la demanda. Un enfoque didáctico y conceptual. Ediciones CEU (Centro Europeo de Estudios Superiores). ISBN.84-88881-22-3.

(2) Valencia Delfa, J. L.; Díaz-Llanos Sainz-Calleja, Fco. J.; Tarazona Lafarga, J. V. (2007). Efecto y utilidad del coeficiente RV de Yves Escoufier en el análisis de correspondencias múltiples. Anales de la Real Academia de Doctores de España.

(3) Foucart, Th. (1997). L'analyse des données. Mode d'emploi. Méthodes et études de cas. Presses Universitaires de Rennes.

(4) Diday, E.; Lemaire, J.; Pouget, J.; Testu, J. (1982). Éléments d'analyse des données. Dunod.

(5) Caillez, P.; Pages, J-P. (1976). Introduction à l'analyse des données. SMASH.

(6) Kendall, M. G.; Stuart, A. (1958-1966). The advanced theory of statistics (3 tomes). London-Griffin.