Anales de la Real Academia de Doctores de España
Volumen 11, pp. 9-44, 2007

EFECTO Y UTILIDAD DEL COEFICIENTE RV DE YVES ESCOUFIER EN EL ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLES

JOSÉ LUIS VALENCIA DELFA
FCO. JAVIER DÍAZ-LLANOS SAINZ-CALLEJA
JOSÉ VICENTE TARAZONA LAFARGA

RESUMEN

El objetivo de este artículo es el de mostrar –de la forma más didáctica posible– el **concepto** y **la utilidad** del **coeficiente RV** en el **análisis de correspondencias múltiples** cuando las **matrices de variables indicadoras** –asociadas a cada una de las **variables cualitativas**– sean de distintas dimensiones. Dicha utilidad, proporciona una **ponderación óptima**,a: **las matrices de variables indicadoras** –asociadas a las **variables cualitativas** de diferente **número de modalidades**– que se encuentran contenidas en una **tabla de datos** sometida a un **ACM.** Dicho de otra manera, es equivalente al efecto de **centrar** y **reducir** una **tabla de datos cuantitativos** con el fin de ser fieles al **principio de homogeneidad de una tabla de datos.** En este sentido S. Wold (1) y J. P. Benzécri (2) proponen transformar las **variables cuantitativas** explicativas de un **análisis discriminante lineal** en los **ejes factoriales** asociados a un **análisis en componentes principales;** es decir, los que aportan la máxima información.

Maurice Roux (3) nos indica que, el **número de ejes factoriales** oscila entre 5 y 10. De esta manera, la **tabla de datos constituida por las variables explicativas, es más homogénea que la tabla de datos de partida.** De esta forma, se consigue que se respete el principio de **homogeneidad.**

Finalmente, mostraremos un ejercicio **exclusivamente didáctico.** Para la ejecución de este ejercicio haremos uso del **AFC** contenido en el paquete de programas STATITCF (Institut Technique des Seréales et des Fourrages).

PABLABRAS CLAVES

Triple estadístico, operador asociado a un triplete estadístico: WD, coeficiente RV, T cuadrado de Tschuprow, análisis de correspondencias simples (AFC), análisis de correspondendias múltiples(ACM), análisis en componentes principales (ACP).

BREVE INTRODUCCIÓN HISTÓRICA DEL COEFICIENTE RV

En 1970, Yves Escoufier(4) introdujo los principios del **coeficiente RV** implementándolos formalmente en 1973(5). Desde entonces, el **coeficiente RV** ha venido siendo, el protagonista de numerosas tesis doctorales,artículos y notas técnicas. A continuación vamos a citar aquellas que –a nuestro juicio– consideramos las más significativas.

En 1976, Yves Escoufier junto con Pierre Robert integraron el **coeficiente RV** en el **Análisis de Datos Multidimensionales Lineal** (6).

Tres años más tarde, en 1979, por un lado,Yves Escoufier impartió un curso de **Análisis de Datos,** profundizando en dicha integración (7); y por otro, junto con Pierre Robert presentó el el método de **elección de variables bajo un contexto de un (ACP),** (8).En ese mismo año, Claude Bernard DO-CHI presentaba su tesis doctoral sobre el tema de **elección de variables** así como el programa informático (9).En 1981, Pierre-Louis Gonzalez (10) y Robert Sabatier (11) aplicaron el **coeficiente RV.** Mientras que en el año 1982, Pierre-Louis Gonzalez presentaba su tesis doctoral sobre el tema de **elección de variables,** aplicandolo a datos **físico-químicos** y **sensoriales** en el café (12). Robert Sabatier, en 1983,presentaba la suya,implementando en ella –además del método de **elección de variables** y el **análisis en componentes principales con respecto a variables instrumentales(ACPVI)–** el **programa informático**, incluyendo además, una aplicación a la **reconstitución de paleoclimas** (13).

En 1985, Fco. Javier Díaz-Llanos introdujo en su tesis doctoral (14), además de un conjunto de **métodos fáctoriales** y **algoritmos de clasificación,** un **test gráfico,** para la retención de variables de naturaleza socioeconómica durante el periodo 1964 a 1979.Un análisis detallado de la justificación, significado e implicaciones del **test gráfico** se encuentra en (14,15). Dicho **test gráfico** propuesto por Yves Escoufier está fuera de hipótesis distribucionales como la Ley de Laplace-Gauss Multidimensional que en algunos casos raros puede verificarse.

En 1986, Yves Escoufier retoma el **ACPVI**- introducido por R.Rao (16) en 1965- bajo la óptica del **coeficiente RV.** (17).

Desde el 1986, hasta nuestros días se han estado realizando numerosos artículos y tesis relevantes sobre el **coeficiente RV.**

En última instancia, hemos de recordar que el **coeficiente RV** –además de haber sido el protaginista del método de **elección de variables bajo el contexto de un ACP–** ha intervenido en el problema de la **detección de individuos atípicos,** ya estudiado bajo el contexto del **Análisis Estadístico Multidimensional Lineal** (en 1963 por S.S. Wilkks (18) y en el 1990 por los investigadores R. Cleroux.; J-M. Helbling y N. Ranger (19), bajo el contexto del **coeficiente RV.** En 1993, F. Crettaz de Roten (20), utilizó el **coeficiente RV** para el estudio de **datos ausentes.**

Breve revisión histórica del análisis de correspondendias simples(AFC) y del análisis de correspondencias múltiples (ACM).

En 1935, Hirschfeld (21) introdujo los principios teóricos del **análisis de correspondencias simples.**Cinco años más tarde Fisher trató dicho tema (22). En 1963, se creó en Rennespor iniciativa de Doyan.Y.Martin - un laboratorio de cálculo equipado con un 1620 IBM. Dos años más tarde, Brigitte Cordier (más tarde Brigitte Escofier), escribía –rápidamente– el primer **análisis de correspondencias** y más tarde, escribía otro, mostrando la relación entre las nubes de puntos N(I) y N (J) (23), siendo esta la originalidad de la escuela francesa.

El primer programa informático del **AFC** fue creado por Brigitte Escofier en el año 1965 (23).

Pero fue J-P Bénzecri y sus colaboradores quienes en los años 70 desarrollaron dicho tema e introdujeron una síntesis del mismo (24).

Aunque, los principios del **análisis de correspondencias múltiples** han sido definidos en 1941 por Guttman (25), seguido en 1950 por Burt (26), y finalmente en 1958 por Hayashi (27); el nombre de **análisis de correspondencias múltiples**, figura –por primera vez– en 1975, por Lebart (28).

En 1988,Brigitte Escofier y Jerôme Pagès exponen didácticamente estos dos **métodos factoriales** en (29).

INTRODUCCIÓN

El hecho de que nos hayamos decidido a realizar este artículo, de la formas **más didáctica posible**, no ha sido por simple capricho sino porque, han pasado **más de 35 años** desde que Yves Escoufier introdujo los principios del **coeficiente RV** (4,5) y ,que nosotros sepamos, aún no se contemplan en ningún libro de texto de Análisis Estadístico Multidimensional Lineal, escrito en español.

Es un hecho sorprendente que, tanto el **coeficiente de correlación lineal de Bravais-Pearson al cuadrado** como la **T cuadrado de Tschuprow**, siendo casos particulares del **coeficiente RV** éstos si se encuentran en los libros de texto de Estadística Descriptiva escritos en español mientras que el **coeficiente RV** no.

El **coeficiente RV** juega un papel importante en el **tratamiento estadístico de las encuestas de opinión.** En éste sentido, hemos de destacar el trabajo realizado por Yves Escoufier, Dambroise y Masotte (30).

El **eje central de éste artículo** es:

Mostrar –de la forma **más didáctica posible–** la aplicación del **coeficiente RV,** en el **análisis de correspondencias múltiples** con el fin de, **proporcionar una ponderación óptima de las matrices de variables indicadoras-** asociadas a las **variables cualitativas** de diferente **número de modalidades-**, contenidas en la **tabla de datos disyuntiva completa,** para ser sometida a un **ACM.**

Cuando se aplica un **ACM** a la **tabla de datos disyuntiva completa**, es aconsejable que el **número de modalidades**- asociado a cada una de las **variables cualitativas**-, no sólo sea el mismo sino también, que sea igual a 4. En el caso particular que el **número de modalidades**- asociado a cada una de las **variables cualitativas**- fuera el mismo, la **transformación** que vamos a realizar a dichas matrices, carece de sentido.

En este artículo, mostraremos, de la forma **más didáctica posible**,una de las múltiples aplicaciones del **coeficiente RV** en el **análisis de correspondencias múltiples.** Se trata de, cómo afecta el primer vector propio ortonormado de la matriz **T cuadrado de Tschuprow,** así como, el **número de modalidades** de cada **variable cualitativa**, a la hora de **proporcionar el efecto de la ponderación óptima de las matrices de variables indicadoras de distintas dimensiones- en cuanto al número de modalidades-** y,la utilidad que representa para la ayuda a la interpretación de los datos empíricos.

Finalmente, mostraremos una aplicación **exclusivamente didáctica** que contenga los conceptos ausentes en los libros de texto de Análisis Estadístico Multidimensional- escritos en español- con el fín de que, se apliquen en los nuevos tratamientos Estadísticos escritos en dicho idioma.

METODOLOGÍA

El proceso metodológico para llevar a cabo los objetivos de este artículo es:

1. Construcción de la **tabla disyuntiva completa.**

2. Construcción de la **tabla de Burt.**

3. Definiciones de forma general de los conceptos de:

    3.1. **Triplete estadístico**

    3.2. **Operador asociado a un triplete estadístico.**

    3.3. El **coeficiente RV.**

4. Particularización de los conceptos ya definidos cuando:

    4.1. Las matrices de datos son las **matrices de variables indicadoras,** asociadas a las **variables cualitativas.**

    4.2. Las matrices de datos son las **matrices de variables indicadoras –centradas por columnas–** asociadas a las **variables cualitativas.**

5. Cálculo de las k(k-1)/2 **T cuadrado de Tschuprow** a partir de las k(k-1)/2 **tablas de contingencia** contenidas en la **tabla de Burt.**

    k: representa el **número de variables cualitativas.**

6. Representación gráfica de las **variables cualitativas.**

7. Análisis de correspondencias múltiples(**ACM**) haciendo uso del **AFC.**

    7.1. Igual **número de modalidades para cada variable cualitativa.**

    7.2. Distinto **número de modalidades** para cada **variable cualitativa.**

1. CONSTRUCCIÓN DE LA TABLA DE DISYUNTIVA COMPLETA

Sean I1,I2,....,IK **variables cualitativas** a m1,m2,....,mk **modalidades**, respectivamente.

A cada una de las **variables cualitativas I1,I2,....,IK** a m1,m2,....,mk **modalidades** le asociamos la matriz

$$U_1,\ U_2, \ldots, U_k$$

constituida por m1, m2,....,mk **variables indicadoras** de dimensiones (n,m1), (n,m2),...., (n,mk), respectivamente.

$$U_1 = \left(I_{11} \ \ I_{12} \cdot\cdot I_{1_{m1}} \right)$$

$$U_2 = \left(I_{21} \ \ I_{22} \cdot\cdot I_{2_{m2}} \right)$$

....

$$U_k = \left(I_{k1} \ \ I_{k2} \cdot\cdot I_{k_{mk}} \right)$$

donde,

$$\left(I_{11} \ \ I_{12} \cdot\cdot I_{1_{m1}} \right)$$

representan las **matrices de variables indicadoras,** asociada a la **variable cualitativa: I1.**

$$I_{21} \ \ I_{22} \ \cdots I_{2_{m2}}$$

representan las **matrices de variables indicadoras,** asociada a la **variable cualitativa: I2.**

$$I_{k1} \ \ I_{k2} \ \cdots I_{k_{mk}}$$

representan las **matrices de variables indicadoras,** asociada a la **variable cualitativa: IK.**

1. CONSTRUCCIÓN DE LA TABLA DISYUNTIVA COMPLETA

A partir de este momento estamos en condiciones de construir la **tabla disyuntiva completa** que se define como la **yustaposición vertical de las matrices de variables indicadoras**, asociadasa las **variables cualitativas**.

$$U = \left(U_1 \ / \ U_2 \ / \ldots / \ U_k \right)$$

2. CONSTRUCCIÓN DE LA TABLA DE BURT

La construcción de la **tabla de Burt** se hace a partir de la **tabla disyuntiva completa** actuando de la siguiente manera:

$$B = U^T U = \begin{pmatrix} U_1^T U_1 & U_1^T U_2 \ldots U_1^T U_k \\ U_2^T U_1 & U_2^T U_2 \ldots U_2^T U_k \\ \ldots\ldots & \ldots\ldots \\ \\ U_k^T U_1 & U_k^T U_2 \ldots U_k^T U_k \end{pmatrix}$$

Hemos de recordar que las k(k-1)/2 **subtablas de Burt** son **tablas de contingencia.**

3. DEFINICIONES DE FORMA GENERAL DE LOS CONCEPTO DE

3.1. Triplete estadístico

En primer lugar, vamos a considerar un **triplete estadístico** y en segundo lugar dos **tripletes estadísticos.**

1. Un **triplete estadístico** está constituido por la siguiente terna de elementos:

$$\left(X, Q, D \right)$$

donde:

X: es la matriz de datos cuantitativos centrada por columnas de dimensiones (n,p)

Q es la métrica introducida en el espacio de los individuos de dimensiones (p,p)

D es la métrica introducica en el espacio de las variables de dimensiones (n,n).

2. Dos **tripletes estadísticos** definidos de la siguiente manera:

$$\left(X, Q_1, D \right) e \left(Y, Q_2, D \right)$$

donde:

X e Y: son las matrices de datos cuantitativos centrados por columnas de dimensiones ((n,p) y (n,q)), respectivamente.

Q1 y Q2 : son las **métricas introducidas en el espacio de los individuos** de dimensiones (p,p) y (q,q), respectivamente.

D : es la **métrica indroducida en el espacio de las variables** de dimensiones (n,n).

3.2.1. *Operador asociado a un triplete estadístico*

A partir del **triplete estadístico**

$$\left(X,\, Q,\, D \right)$$

le asociamos el **operador WD** definidode la siguiente manera:

$$WD \;=\; X\,Q\;X^{T}\;D$$

3.2.2. *Operadores asociados a dos tripletes estadísticos.*

$$\left(X\,,\; Q_{1},\, D \right)\; y\; \left(Y\,,\; Q_{2},\, D \right)$$

A estos dos **tripletes estadísticos** les asociamos los **operadores**

$$W_{1}D \quad y \quad W_{2}D$$

definidos de la siguiente manera:

$$W_{1}D = X\,Q_{1}\;\; X^{T}\; D\;\; y\;\; W_{2}D = Y\,Q_{2}\;\; Y^{T}\; D$$

3.3. Coeficiente RV

3.3.1. *Definición*

El **coeficiente RV** entre **operadores** se define de la siguiente manera:

$$RV\left(\; W_{1}D, W_{2}D \;\right) = \frac{Tr\left(\; W_{1}D \;\; W_{2}D \;\right)}{\sqrt{Tr\;\left(\; W_{1}D \;\right)^{2}\; Tr\;\left(\; W_{2}D \;\right)^{2}}}$$

3.3.2. *Propiedades*

$$1^{a} \quad 0 \;\leq\; RV\left(\; W_{1}D, W_{2}D \;\right) \;\leq\; 1$$

$$2^{a} \quad RV\left(\; W_{1}D,\;\; W_{2}D \;\right) = 1 \;\; ssi \exists k \;\varepsilon\; R_{+}^{*} \quad / \quad W_{1}D = \; k\; W_{2}D$$

$$RV\left(\; W_{1}D,\;\; W_{2}D \;\right) = 0 \quad ssi S_{12} = 0$$

4. PARTICULARIZACIÓN DE LOS CONCEPTOS YA DEFINIDOS CUANDO

4.1. Las matrices de datos son las matrices de variables indicadoras asociadas a las variables cualitativas.

Sustitución de

X por U_1

Y por U_2

donde U_1 y U_2 son las matrices

de variables indicadoras de dimensiones (n,m_1) y (n,m_2)

asociadas a la primera y a la segunda variable cualitativa

La matriz $P = U_1^T\ D\ U_2$ de dimensiones (m_1,m_2)

y cuyos elementos estan caracterizados

por $p(i1,i2) = \dfrac{n(i1,i2)}{n}$ la definimos de la siguiente manera:

Por consiguiente,

$$P = \begin{pmatrix} p_{11}\cdots\cdots p_{12}\cdots\cdots p_{1m_2} \\ p_{21}\cdots\cdots p_{22}\cdots\cdots p_{2m_2} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ p_{m_{11}}\cdots\cdot p_{m_{12}}\cdots\cdot p_{m_1,m_2} \end{pmatrix}$$

Las métricas Q_1 y Q_2 van ser

D_{I1}^{-1} y D_{I2}^{-1}, respectivamente.

La definición de dichas nuevas métricas es la siguiente:

$$D_{I1}^{-1} = diag\left(\frac{1}{p_{1+}}, \frac{1}{p_{2+}}, \ldots\ldots, \frac{1}{p_{m_1+}} \right)$$

$$D_{I2}^{-1} = diag\left(\frac{1}{p_{+1}}, \frac{1}{p_{+2}}, \ldots\ldots, \frac{1}{p_{+m_2}} \right)$$

$$D = \frac{1}{n}\ I_{nxn}$$

donde,

I_{nxn} es la matriz identidad de orden n

4.1.1. *Tripletes estadísticos*

Los **tripletes estadísticos** adoptan la siguiente forma

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

4.1.2. *Operadores WD*

A cada uno de los **tripletes estadísticos** asociamos sus **operadores WD** de la siguiente manera.

$$\left(U_1 \ , \ D_{I1}^{-1}, D \right) \dashrightarrow W_1 D = U_1 \ D_{I1}^{-1} \ U_1^T \ D$$

$$\left(U_2 \ , \ D_{I2}^{-1}, D \right) \dashrightarrow W_2 D = U_2 \ D_{I2}^{-1} \ U_2^T \ D$$

4.1.3. *Coeficiente RV*

Para evaluar el cálculo del **coeficiente RV** es necesario calcular las siguientes trazas:

$$Tr\left(W_1 D \right)^2, Tr\left(W_2 D \right)^2 \ y \ Tr\left(W_1 D \ W_2 D \right)$$

1º *Cálculo de*: $Tr(W_1 D)^2$

Teniendo en cuenta:

1.º La propiedad conmutativa de la traza de un producto de matrices.

$$2^{\text{o}} \ U_1^T \ D \ U_1 = D_{I1}$$

$$3^{\text{o}} \ D_{I1}^{-1} \ D_{I1} = I_{m_1 \cdot m_1}$$

donde,

$I_{m_1 \cdot m_1}$ *es la matriz identidad de orden* m_1

tenemos,

$$Tr\left(W_1 D \right)^2 = Tr\left(U_1 \ D_{I1}^{-1} \ U_1^T \ D \ U_1 \ D_{I1}^{-1} \ U_1^T \ D \right) =$$

$$= Tr\left(D_{I1}^{-1} \ U_1^T \ D \ U_1 \ D_{I1}^{-1} \ U_1^T \ D \ U_1 \right) =$$

$$= Tr\left(D_{I1}^{-1} \ D_{I1} \ D_{I1}^{-1} \ D_{I1} \right) = Tr\left(I_{m_1 \cdot m_1} \right) = m_1$$

Por tanto,

$$Tr\left(W_1 D \right)^2 = m_1$$

2º *Cálculo de:* $Tr\left(W_2 D \right)^2$

Teniendo en cuenta:

1.º La propiedad conmutativa de la traza de un producto de matrices.

$$2^{\text{o}} \ U_2^T \ D \ U_2 = D_{I2}$$

$$3^{\text{o}} \ D_{I2}^{-1} \ D_{I2} = I_{m_2 \cdot m_2}$$

donde,

$I_{m_2 \cdot m_2}$ *es la matriz identidad de orden* m_2

tenemos,

$$Tr\left(W_2 D \right)^2 = Tr\left(U_2 \ D_{I2}^{-1} \ U_2^T \ D \ U_2 \ D_{I2}^{-1} \ U_2^T \ D \right) =$$

$$= Tr\left(D_{I2}^{-1} \ U_2^T \ D \ U_2 \ D_{I2}^{-1} \ U_2^T \ D \ U_2 \right) =$$

$$= Tr\left(D_{I2}^{-1} \ D_{I2} \ D_{I2}^{-1} \ D_{I2} \right) = Tr\left(I_{m_2 \cdot m_2} \right) = m_2$$

Por tanto,

$$Tr\left(W_2 D\right)^2 = m_2$$

$$3^{\underline{o}}\ \textit{Cálculo de: } Tr\left(W_1 D\ W_2 D\right)$$

Teniendo en cuenta:

1.º La propiedad conmutativa de la traza de un producto de matrices:

$$2^{\underline{o}}\ U_1^T D\ U_2 = P$$

y operando convenientemente tenemos,

$$Tr\left(W_1 D\ W_2 D\right) = Tr\left(U_1\ D_{I1}^{-1}\ U_1^T\ D\ U_2\ D_{I2}^{-1}\ U_2^T\ D\right) =$$
$$= Tr\left(D_{I1}^{-1}\ U_1^T\ D\ U_2\ D_{I2}^{-1}\ U_2^T\ D\ U_1\right) =$$

$$= Tr\left(D_{I1}^{-\frac{1}{2}}\ D_{I1}^{-\frac{1}{2}}\ U_1^T\ D\ U_2\ D_{I2}^{-\frac{1}{2}}\ D_{I2}^{-\frac{1}{2}}\ U_2^T\ D\ U_1\ D_{I1}^{-\frac{1}{2}}\right) =$$

$$= Tr\left[\left(D_{I1}^{-\frac{1}{2}}\ U_1^T\ D\ U_2\ D_{I2}^{-\frac{1}{2}}\right)\left(D_{I1}^{-\frac{1}{2}}\ U_1^T\ D\ U_2\ D_{I2}^{-\frac{1}{2}}\right)^T\right] =$$

$$= Tr\left[\left(D_{I1}^{-\frac{1}{2}}\ P\ D_{I2}^{-\frac{1}{2}}\right)\left(D_{I1}^{-\frac{1}{2}}\ P\ D_{I2}^{-\frac{1}{2}}\right)^T\right] =$$

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

Por tanto,

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

Acto seguido, estamos interesados en encontrar una relación entre,

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

Desarrollando convenientemente la expresión,

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

llegamos al siguiente resultado,

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

tal como mostramos a continuación,

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

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

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

Teniendo en cuenta los siguientes resultados,

$$\sum_{i1=1}^{i1=m_1}\sum_{i2=1}^{i2=m_2}p(i1,i2)=1\qquad\sum_{i1=1}^{i1=m_1}p_{i1+}=1\qquad\sum_{i2=1}^{i2=m_2}p_{+i2}=1$$

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

De lo que se desprende que,

$$Tr\left(\ W_1D\ W_2D\ \right)=\sum_{i1=1}^{i1=m_1}\sum_{i2=1}^{i2=m_2}\frac{\left[p(i1,i2)-\ p_{i1+}\ p_{+i2}\right)^2}{p_{i1+}\ p_{+i2}}\ +\ 1$$

Una vez definidas y calculadas las trazas asociadas a los **operadores,**

$$Tr\left(\ W_1D\ \right)^2,\ Tr\left(\ W_2D\ \right)^2\ y\ Tr\left(\ W_1D\ W_2D\ \right)$$

estamos en condiciones de calcular el **coeficiente RV**

COEFICIENTE RV

Sustituyendo los resultados obtenidos –referentes a las trazas de los operadores– en la fórmula del **coeficiente RV**, tenemos que:

$$RV\left(\ W_1D,\ W_2D\ \right)=\frac{\displaystyle\sum_{i1=1}^{i1=m_1}\sum_{i2=1}^{i2=m_2}\frac{\left[p(i1,i2)-\ p_{i1+}\ p_{+i2}\right]^2}{p_{i1+}\ p_{+i2}}\ +\ 1}{\sqrt{m_1m_2}}$$

4.2. Las matrices de datos son las **matrices de variables indicadoras –centradas por columnas–** asociadas a las **variables cualitativas.**

En esta situación, vamos a proceder a las **transformaciones de las matrices de variables indicadoras**, asociadas a las **variables cualitativas I1 e I2.** El hecho de que hagamos esta **transformación,** no es por puro capricho sino que haciéndola nos desembarazamos del 1 que aparece reflejado en el **coeficiente RV** y así, se verifica que el nuevo **coeficiente RV** entre operadores centrados es igual a la **T cuadrado de Tschuprow,** entre la **variable cualitativa I1** y la **variable cualitativa I2.** A título informativo hemos de indicar que, la **T cuadrado de Tschuprow** ha sido ya utilizada por el profesor Gilbert Saporta en su método de **análisis discriminante lineal cuando, las variables a explicar son cualitativas** (31).

La **transformación de las matrices de variables indicadoras** asociada, a las **variables cualitativas I1 e I2** consiste en, **centrar por columnas,** dichas matrices de la siguiente manera:

$$U_1: \ U_1^* = \left(I_{nxn} - 1_n 1_n^T D \right) U_1$$

$$U_2: \ U_2^* = \left(I_{nxn} - 1_n 1_n^T D \right) U_2$$

donde,

I_{nxn}: es la matriz identidad de orden n.

1_n: es un vector columna que contiene n unos.

1_n^T: es el vector transpuesto de 1_n.

D: es una matriz de orden n definida de

la siguiente manera: $D = \frac{1}{n} I_{nxn}$

A partir de este momento, vamos a actuar de la misma manera que la mostrada con anterioridad pero, en lugar de considerar las **matrices de variables indicadoras,** asociadas a las **variables cualitativas I1 e I2,** trabajaremos, con las **matrices de variables indicadoras –centradas por columnas–.**

U_1^* *y* U_2^*

Sustitución de:

X por U_1^*

Y por U_2^*

donde,

U_1^* y U_2^* son las matrices

de variables indicadoras centradas por columnas

de dimensiones (n,m_1) y (n,m_2) de la

primera y segunda variable cualitativa.

C es la matriz $U_1^{*T} D U_2^*$

de dimensiones(m_1,m_2),cuyo elemento genérico es:

$$C(i1,i2) = p(i1,i2) - p_{i1+} \ p_{+i2}$$

Las métricas Q_1, Q_2 y D

van a ser las mismas que en al **primera situación.**

4.2.1. *Tripletes estadísticos*

En nuestro caso concreto, los dos **tripletes estadísticos,** están contituidos por las **matrices de variables indicadoras –centradas por columnas–** asociadas a las **variables cualitativas I1 e I2**, la **métrica** asociada **al espacio de los individuos** y, la **métrica** asociada al **espacio de las variables indicadoras.**

Por tanto, los dos **tripletes estadísticos** en nuestro caso concreto son los siguientes:

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

4.2.2. *Operadores asociados a los dos tripletes estadísticos*

A cada **triplete estadístico** asociamos su **operador** de la siguiente manera:

$$\left(U_1^*,\ D_{I1}^{-1},\ D \right) \dashrightarrow W_1^* D = U_1^*\ D_{I1}^{-1}\ U_1^{*T}\ D$$

$$\left(U_2^*,\ D_{I2}^{-1},\ D \right) \dashrightarrow W_2^* D = U_2^*\ D_{I2}^{-1}\ U_2^{*T}\ D$$

4.2.3. *Coeficiente RV*

Para la evaluación del **coeficiente RV** es necesario calcular las siguientes trazas:

$$Tr\left(W_1^* D \right)^2,\ Tr\left(W_2^* D \right)^2 \ y\ Tr\left(W_1^* D\ W_2^* D \right)$$

$$1^{\underline{o}} : \textit{Cálculo de: } Tr\left(W_1^* D \right)^2$$

Teniendo en cuenta:

1.º La propiedad conmutativa de la traza del producto de matrices.

$$2^{\underline{o}} \quad U_I^T\ D\ U_I = D_{II}$$

$$3^{\underline{o}} \quad U_I^T\ D\ 1_n = D_{II}\ 1_{II}$$

$$4^{\underline{o}} \quad U_I\ 1_{II} = 1_n$$

donde,

1_{II} *es un vector columna que contiene* m_I *unos*

y operando convenientemente tenemos,

$$Tr\left(W_1^* D \right)^2 = Tr\left(U_I^*\ D_{II}^{-1}\ U_I^{*T}\ D\ U_I^*\ D_{II}^{-1}\ U_I^{*T}\ D \right) =$$

$$= Tr\left(D_{II}^{-1}\ U_I^{*T}\ D\ U_I^*\ D_{II}^{-1}\ U_I^{*T}\ D\ U_I^* \right)$$

Antes de seguir adelante es conveniente expresar

$$U_I^{*T}\ D\ U_I^* \textit{ en función de } D_{II} \textit{ y de } 1_{II}$$

Para conseguir este objetivo operamos de la siguiente manera

$$U_1^{*T}\ D\ U_1^{*} = U_1^{T}\left(\ I_{nxn} - D\ 1_n 1_n^{T}\ \right)^{T} D\left(I_{nxn} - D\ 1_n 1_n^{T}\ \right) U_1 =$$

$$= U_1^{T}\ D\ U_1 - U_1^{T}\ D\ 1_n 1_n^{T}\ D\ U_1 - U_1^{T}\ D\ 1_n 1_n^{T}\ D\ U_1 +$$

$$+ U_1^{T}\ D\ 1_n\ 1_n^{T}\ D\ 1_n 1_n^{T}\ D\ U_1 =$$

$$= U_1^{T}\ D\ U_1 - U_1^{T}\ D\ 1_n\ 1_n^{T}\ D\ U_1 = D_{11}\ - D_{11}\ 1_{11} 1_{11}^{T}\ D_{11}$$

Por tanto,

$$Tr\left(\ D_{11}^{-1}\ U_1^{*T}\ D\ U_1^{*}\ D_{11}^{-1}\ U_1^{*T}\ D\ U_1^{*}\ \right) =$$

$$= Tr\left[\ D_{11}^{-1}\left(D_{11} -\ D_{11}\ 1_{11} 1_{11}^{T} D_{11}\right) D_{11}^{-1}\left(D_{11} - D_{11}\ 1_{11} 1_{11}^{T} D_{11}\right)\right] =$$

$$= Tr\left[\left(I_{m_1,m_1} -\ 1_{11} 1_{11}^{T}\right)\left(I_{m_1,m_1}\ - 1_{11} 1_{11}^{T} D_{11}\right)\right] =$$

$$= Tr\left(I_{m_1,m_1}\right) - Tr\left(\ 1_{11} 1_{11}^{T}\ D_{11}\ \right) - Tr\left(1_{11} 1_{11}^{T}\ D_{11}\right) +$$

$$+ Tr\left(1_{11} 1_{11}^{T}\ D_{11}\ 1_{11} 1_{11}^{T}\ D_{11}\right) =$$

$$= Tr\left(I_{m_1,m_1}\right) - Tr\left(\ 1_{11} 1_{11}^{T}\ D_{11}\ \right) =$$

$$= m_1\ - Tr\left(\ 1_{11} 1_{11}^{T}\ D_{11}\right) = m_1\ -1$$

De lo que se desprende que:

$$Tr\left(\ W_1^{*} D\ \right)^2 = m_1 - 1$$

$$2^{\underline{o}}: \textit{Cálculo de: } Tr\left(\ W_2^{*} D\ \right)^2$$

Teniendo en cuenta:

1. La propiedad conmutativa de la traza del producto de matrices

$$2^{\underline{a}}\ U_2^{T}\ D\ U_2 = D_{12}$$
$$3^{\underline{a}}\ U_2^{T}\ D\ 1_n = D_{12}\ 1_{12}$$
$$4^{\underline{a}}\ U_2\ 1_{12} = 1_n$$

donde,

$$1_{I2} \textit{ es un vector columna que contiene } m_2 \textit{ unos}$$

y operando convenientemente tenemos,

$$Tr\left(W_2^* D \right)^2 = Tr\left(U_2^* \; D_{I2}^{-1} \; U_2^{*T} \; D \; U_2^* \; D_{I2}^{-1} \; U_2^{*T} \; D \right) =$$

$$= Tr\left(D_{I2}^{-1} \; U_2^{*T} \; D \; U_2^* \; D_{I2}^{-1} \; U_2^{*T} \; D \; U_2^* \right)$$

Antes de seguir adelante es conveniente expresar

$$U_2^{*T} \; D \; U_2^* \textit{ en función de } D_{I2} \textit{ y } 1_{I2}$$

Para ello operamos de la siguiente manera,

$$U_2^{*T} \; D \; U_2^* \; = \; U_2^T (I_{nxn} - D \, 1_n 1_n^T)^T \; D \left(I_{nxn} - 1_n 1_n^T \; D \right) U_2 =$$

$$= U_2^T \; D \; U_2 - U_2^T \; D \; 1_n 1_n^T \; D \; U_2 - U_2^T \; D \; 1_n 1_n^T U_2 +$$

$$+ \; U_2^T \; D \; 1_n 1_n^T \; D \; U_2 = U_2^T \; D \; U_2 - U_2^T \; D \; 1_n 1_n^T \; D \; U_2$$

De lo que se desprende,

$$U_2^{*T} \; D \; U_2^* \; = \; D_{I2} \; - \; D_{I2} \; 1_{I2} 1_{I2}^T \; D_{I2}$$

Por consiguiente,

$$Tr\left(D_{I2}^{-1} \; U_2^{*T} \; D \; U_2^* \; D_{I2}^{-1} \; U_2^{*T} \; D \; U_2^* \right) =$$

$$= Tr\left[\left(D_{I2}^{-1}\left(D_{I2} - D_{I2} 1_{I2} 1_{I2}^T \; D_{I2}\right) D_{I2}^{-1} (D_{I2} - D_{I2} 1_{I2} 1_{I2}^T D_{I2} \right)\right] =$$

$$= Tr\left[\left(\; I_{m2 \cdot m2} - 1_{I2} 1_{I2}^T D_{I2})(I_{m2 \cdot m2} - 1_{I2} 1_{I2}^T \; D_{I2} \right)\right] =$$

$$= Tr\left(I_{m2 \cdot m2} \right) - Tr\left(1_{I2} 1_{I2}^T \; D_{I2} \right) = \; m_2 - Tr\left(1_{I2} 1_{I2}^T D_{I2} \right) = \; m_2 \; -1$$

De lo que se desprende,

$$Tr\left(W_2^* D \right)^2 \; = \; m_2 \; - \; 1$$

$$3^{\underline{o}} \textit{ Cálculo de : } Tr\left(W_1^* D \; W_2^* D \right)$$

Teniendo en cuenta

1. La propiedad conmutativa de la traza de matrices y operando convenientemente tenemos,

$$Tr\left(W_1^* D\ W_2^* D \right) = Tr\left(U_1^*\ D_{I1}^{-1}\ U_1^{*T}\ D\ U_2^*\ D_{I2}^{-1}\ U_2^{*T}\ D \right) =$$
$$= Tr\left(D_{I1}^{-1}\ U_1^{*T}\ D\ U_2^*\ D_{I2}^{-1}\ U_2^{*T}\ D\ U_1^* \right)$$

Antes de seguir adelante es conveniente expresar

$U_1^{*T}\ D\ U_2^*$ *en función de* U_1 *y* U_2

Para ello operamos de la siguiente manera,

$$U_1^{*T}\ D\ U_2^* = U_1^T (I_{nxn} - D\ 1_n 1_n^T)^T\ D\ (I_{nxn} - 1_n 1_n^T\ D) U_2 =$$

$$= U_1^T\ D\ U_2 - U_1^T\ D\ 1_n 1_n^T\ D\ U_2 - U_1^T\ D\ 1_n 1_n^T\ D\ U_2 +$$

$$+ U_1^T\ D\ 1_n\ 1_n^T\ D\ 1_n\ 1_n^T\ D\ U_2 =$$

$$= U_1^T\ D\ U_2 - U_1^T\ D\ 1_n 1_n^T\ D\ U_2$$

Por tanto,

$$U_1^*\ D\ U_2^* = U_1^T\ D\ U_2 - U_1^T\ D\ 1_n 1_n^T\ D\ U_2$$

Sea la matriz $C = U_1^{*T}\ D\ U_2^*$ *de elementos genéricos*

$$C(i1,i2) = p(i1,i2) - p_{i1+}\ p_{+i2}$$

Teniendo en cuenta este resultado se desprende,

$$Tr\left(W_1^* D\ W_2^* D \right) = \sum_{i1=1}^{i1=m_1} \sum_{i2=1}^{i2=m_2} \frac{\left[p(i1,i2) - p_{i1+}\ p_{+i2} \right]^2}{p_{i1+}\ p_{+ç2}}$$

Una vez definidas y calculadas las trazas asociadas a los **operadores,**

$$Tr\left(W_1^* D \right)^2,\ Tr\left(W_2^* D \right)^2\ y\ Tr\left(W_1^* D\ W_2^* D \right)$$

calculamos, sin dificultad, el **coeficiente RV** mediante la siguiente fórmula,

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

La traducción de esta relación en términos de estimación empíricas revela que no es más que la **T cuadrado de Tschuprow** entre **I1 e I2** adoptando la siguiente fórmula,

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

que operando convenientemente llegamos a una fórmula más operativa que la anterior,

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

5. Cálculo de las k(k-1)/2 T cuadrado de Tschuprow a partir de la k(k-1)/2 **tablas de contingencia** contenidas en la **tabla de Burt.**

A partir de este momento estamos en condiciones de evaluar la k(k-1)/2 **T cuadrado de Tschuprow**, y acto seguido, podremos construir la correspondiente matriz.

$$T^2 = \begin{pmatrix} T^2_{I1,I1} & T^2_{I1,I2} \cdots\cdots T^2_{I1,IK} \\ T^2_{I2,I1} & T^2_{I2,I2} \cdots\cdots T^2_{I2,IK} \\ \cdots\cdots & \cdots\cdots\cdots\cdots \\ T^2_{IK,I1} & T^2_{IK,I2} \cdots\cdots T^2_{IK,IK} \end{pmatrix}$$

A título informativo hemos de indicar que, aunque la **T cuadrado de Tschuprow**, es un estadístico conocido desde hace tiempo y- que nosotros sepamos- contemplado por primera vez en (1958-1966) por Kendall y Stuart (32). Más tarde ha sido introducido, en 1976 por Caillez y Pages (33), en 1979 por Escoufier (7), en 1984 por Foucart (34), en 1987 por Matín Pliego (35), en 1988 por García Mouton (36), en 1989 por García Moutón y Díaz-Llanos (37), en 1994 por Celeux y Nakache (38), en 1995 por Díaz-Llanos (39) y en 1999 por Saporta (40).

Observación de interés: En 1978, J-J-Daudin (41) demuestra que, la nulidad de la **T de Tschuprow parcial,** introducida por Saporta y comentada por Celeux y Nakache (38), no implica la **independencia condicional.**

6. Representación gráfica de las variables cualitativas I1,I2,...,IK en los círculos de correlacción.

La diagonalización de la matriz

$$T^2$$

(**matriz simétrica semi-definida positiva**) nos permite la representación de las **variables cualitativas I1,I2,....IK** en los **círculos de correlaciones.** De tal manera que, las coordenadas de las **variables cualitativas I1,I2,...,IK,** en los **círculos de correlaciónes,** se calculan de la siguiente manera.

$$I1 \quad L_{11}\sqrt{\lambda_1} \quad L_{21}\sqrt{\lambda_2}\ldots\ldots L_{K1}\sqrt{\lambda_K}$$

$$I2 \quad L_{12}\sqrt{\lambda_1} \quad L_{22}\sqrt{\lambda_2}\ldots\ldots L_{K2}\sqrt{\lambda_K}$$

....

$$IK \quad L_{1K}\sqrt{\lambda_1} \quad L_{2K}\sqrt{\lambda_2}\ldots\ldots L_{KK}\sqrt{\lambda_K}$$

Nota: aconsejamos muy vivamente que el cálculo de los **valores propios**

$$(\lambda_1 > \lambda_2 > \ldots > \lambda_K)$$

se realice aplicando el método de Givens-Householder y el de los **vectores propios** haciendo uso del método Q-L implícito.

Los **vectores propios** asociados a los **valores propios** que vamos a mostrar son **ortonormados**:

$$L_1 = \begin{pmatrix} L_{11} \\ L_{12} \\ . \\ L_{1K} \end{pmatrix} \quad L_2 = \begin{pmatrix} L_{21} \\ L_{22} \\ . \\ L_{2K} \end{pmatrix} \ldots\ldots\ldots L_K = \begin{pmatrix} L_{K1} \\ L_{K2} \\ . \\ L_{KK} \end{pmatrix}$$

Los métodos mencionados- tanto para el cálculo de los **valores propios** como, para el de los **vectores propios**- están contenidos en el libro de Householder (42), los apuntes de cátedra de Puy Huarte (43) y en los libros de Ciarlet (44) y de Ralston (45).

6.1. Representacion gráfica de las variables cualitativas I1 ,I2, ...,IK en los **círculos de correlaciones (1-2), (1-3) y (2-3).**

$$Sean\ \lambda_1, \lambda_2\ y\ \lambda_3$$

los tres primeros **valores propios** de la **matriz de Tschuprow**.

Sean

$$L_1 = \begin{pmatrix} L_{11} \\ L_{12} \\ . \\ L_{1k} \end{pmatrix} \quad L_2 = \begin{pmatrix} L_{21} \\ L_{22} \\ . \\ L_{2K} \end{pmatrix} \quad L_3 = \begin{pmatrix} L_{31} \\ L_{32} \\ . \\ L_{3K} \end{pmatrix}$$

los **vectores propios ortonormados** asociados tres primeros **valores propios**.

6.1.1. *Cálculo de las coordenadas de las variables cualitativas I1, I2,...., IK en los círculos de correlaciones (1-2), (1-3) y (2-3)*

$$I1 \quad L_{11}\sqrt{\lambda_1} \quad L_{21}\sqrt{\lambda_2} \quad L_{31}\sqrt{\lambda_3}$$

$$I2 \quad L_{12}\sqrt{\lambda_1} \quad L_{22}\sqrt{\lambda_2} \quad L_{32}\sqrt{\lambda_3}$$

....

$$IK \quad L_{1K}\sqrt{\lambda_1} \quad L_{2K}\sqrt{\lambda_2} \quad L_{3K}\sqrt{\lambda_3}$$

7. ANÁLISIS DE CORRRESPONDENCIAS MÚLTIPLES (ACM)

Vamos a considerar dos **situaciones**, en cuanto se refiere al **número de modalidades** de las **variables cualitativas.**

Primera situación: El **número de modalidades de las variables cualitativas** es el mismo. **Situación poco** frecuente en un cuestionario de opinión ya que, por lo general, el **número de modalidades** no es el mismo para cada **variable cualitativa.** En esta **situación,** no es necesario hacer uso de ninguna **transformación** a las **matrices de variables indicadoras**, asociadas a las **variables cualitativas.**

Realizaremos el **ACM,** haciendo uso de un **AFC,** disponiendo la **tabla de datos de la siguiente manera:**

$$T_{BU} = \begin{pmatrix} B \\ - \\ U \end{pmatrix}$$

donde,

U es la **tabla disyuntiva completa** de dimensiones(n,km) definida de la siguiente manera,

$$U = \left(U_1 \ / \ U_2 \ / \ldots\ldots / \ U_K \right)$$
$$U_j \quad j = 1,\ldots.,K$$

son las **matrices de variables indicadoras** de dimensiones (n,m), asociadas a cada una de las **variables cualitativas I1,I2,...,IK.**

La **yuxtaposición vertical** de las **matrices de variables indicadoras,** asociadas a las **variables cualitativas**, constituye la **tabla disyuntiva completa.**

n: es el **número de individuos**
m: es el **número de modalidades**
k: es el **número de variables cualitativas**

B es la **tabla de Burt** de dimensiones (km,km) definida de la siguiente manera,

$$B = U^T \; U = \begin{pmatrix} B_{11} \; B_{12} \cdots\cdots B_{1k} \\ B_{21} \; B_{22} \cdots\cdots B_{2K} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \\ B_{K1} \; B_{K2} \cdots\cdots B_{KK} \end{pmatrix}$$

$$B_{ij} = U_i^T \; U_j \quad , i = 1, \ldots\ldots, K \; j = 1, \ldots\ldots, K$$

son los bloques de la **tabla de Burt.**

Las filas de la **tabla de Burt** se consideran como **puntos-individuos activos** y, las filas de la **tabla disyuntiva completa**, se consideran como **puntos-individuos suplementarios.**

Segunda situación: El **número de modalidades** de las **variables cualitativas,** no es el mismo. En esta **situación,** es necesario **transformar** las **matrices de variables indicadoras** asociadas a cada una de las **variables cualitativas.**

Realizaremos el **ACM,** haciendo uso del **AFC,** disponiendo la **Tabla de datos de la siguiente manera,**

$$T_{[BU]^o} = \begin{pmatrix} B^o \\ \text{---} \\ U^o \end{pmatrix}$$

donde,

$$U^o \textit{ es la tabla disyuntiva completa transformada}$$

$$\textit{de dimensiones} \left(n, \; \sum_{j=1}^{j=K} m_j \right)$$

$$\textit{definida de la siguiente manera}$$

$$U^o = \left(\; U_1^o \; / \; U_2^o \; / \ldots\ldots / \; U_K^o \; \right)$$

Los elementos de esta última **tabla de datos** significan:

U_j^o: son las matrices de variables indicadoras transformadas de dimensiones $(n,m_1),(n,m_2),...,(n,m_K)$, respectivamente asociadas a las variables cualitativas $I1, I2,...,IK$, respectivamente.

$$U_j^o = \frac{U_j}{\sqrt{m_j - 1}} L_{1j} \quad ,j = 1,.....,K$$

donde,

$U_j \;\; j = 1,....,K$ son las matrices de variables indicadoras asociadas a las variables cualitativas.

$$L_{1j} \quad ,j = 1,.....,K$$

son las componentes del **primer vector propio ortonormado de la matriz T cuadrado de Tschuprow.**

$$m_j$$

es el **número de modalidades** de las **variables cualitativas. I1, I2,,....,IK.**

B^o es la tabla de Burt transformada de dimensiones $\left(\sum_{j=1}^{j=K} m_j , \sum_{j=1}^{j=K} m_j \right)$

definida de la siguiente manera,

$$B^o = U^{oT} \; U^o = \begin{pmatrix} B_{11}^o & B_{12}^o \cdots\cdots B_{1K}^o \\ B_{21}^o & B_{22}^o \cdots\cdots B_{2K}^o \\ \cdots\cdots & \cdots\cdots\cdots\cdots \\ \\ B_{K1}^o & B_{K2}^o \cdots\cdots B_{KK}^o \end{pmatrix}$$

$$B_{ij}^o = U_i^{oT} \; U_j^o \quad ,i = 1,.....,K \;\; j = 1,.....,K$$

son los bloques de la **tabla de Burt transformada.**

Las filas de la **tabla de Burt –transformada–** se consideran como **puntos-individuos activos** y, las filas de la **tabla disyuntiva completa- transformada-** se consideran como **puntos-individuos suplementarios.**

Este procedimiento que hemos presentado, para la realización de un **ACM** mediante el **AFC,** se encuentra reflejado- brevemente- en 1982 por Diday, Lemaire y Testu (46) y, también se ha propuesto como una estrategia metodológica, para el **cierre de ventas,** en 2002 por Díaz-Llanos (47).

Ejercicio didáctico que muestra los conceptos introducidos en la parte teórica.

Tres variables cualitativas: I1, I2 e I3

I1 es una **variable cualitativa** a dos modalidades I11 e I12 cuyas **variables indicadoras** son:

$$I_{11} \; e \; I_{12}$$

I2 es una **variable cualitativa** a 3 modalidades I21,I22 e I23 cuyas **variables indicadoras** son:

$$I_{21}, \; I_{22} \; e \; I_{23}$$

I3 es una **variable cualitativa** a 4 modalidades I31,I32,I33 e I34 cuyas **variables indicadoras** son:

$$I_{31}, \; I_{32}, I_{33} \; e \; I_{34}$$

Teniendo en cuenta la asignación de la codificación de cada una de las modalidades de I1, I2 e I3, no afecta para el cálculo de los **valores propios de B (tabla de Burt) (anexo 1),** consideramos la que mostramos a continuación

SITUACIÓN		
Variables cualitativas	*Modalidades*	*Codificación*
I1	I11 I12	1 2
I2	I21 I22 I23	1 2 3
I3	I31 I32 I33 I34	1 2 3 4

A partir de esta información, estamos en condiciones de construir una **tabla numérica,** asociada a las tres **variables cualitativas** de una muestra de cuatro individuos, que se caracterizan, según se indica en la **tabla numérica**. Esta **tabla,** no es susceptible de ser sometida a ningún tipo de análisis estadístico, ya que, los **números** que contiene carecen de significado.

TABLA NUMÉRICA

	I1	*I2*	*I3*
1	2	3	2
2	1	2	1
3	2	2	4
4	2	1	3

A partir de la **tabla numérica** podemos construir la **tabla disyuntiva completa**

TABLA DISYUNTIVA COMPLETA

I1		*I2*			*I3*			
I11	*I12*	*I21*	*I22*	*123*	*I31*	*I32*	*I33*	*134*
0	1	0	0	1	0	1	0	0
1	0	0	1	0	1	0	0	0
0	1	0	1	0	0	0	0	1
0	1	1	0	0	0	0	1	0

La **tabla disyuntiva completa** está constituida por, la **yustaposición vertical** de las **matrices de variables indicadoras,** asociadas a cada una de las **variables cuantitativas I1,I2 e I3.**

Dichas **matrices de variables indicadoras** son:

$$U_1 \text{ , } U_2 \text{ y } U_3$$

de dimensiones (4,2), (4,3) y (4,4), respectivamente.

Así pues, en nuestro caso concreto, las **matrices de variables indicadoras,** asociadas a las **variables cualitativas** son:

$$U_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \; U_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \; U_3 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Construcción de la tabla de Burt

$$B = \begin{pmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{pmatrix}$$

$$B_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} \quad B_{12} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \quad B_{13} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{pmatrix}$$

$$B_{21} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} \quad B_{22} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad B_{23} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$B_{31} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \quad B_{32} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad B_{33} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

donde,

$$B_{12}\ ,\ B_{13}\ y\ B_{23}$$

representan las **tablas de contingencias** a partir de las cuales, estamos en condiciones de calcular las tres **T cuadrado de Tschuprow.**

Cálculo de las tres T cuadrado de Tschuprow a partir de las tres **tablas de contingencia** contenidas en la **tabla de Burt**

1º *Cálculo de:* $T^2_{I1,I2}$

Aplicando la fórmula de $T^2_{I1,I2}$

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

a la tabla de contingencia B_{12}

obtenemos el siguiente resultado,

$$T^2_{I1,I2} = \frac{\left(\frac{1}{2} + \frac{1}{3}\,\frac{1}{6} + \frac{1}{3}\right) - 1}{\sqrt{(1)(2)}} = 0{,}2357$$

2º Cálculo de: $T^3_{I1,I3}$

Aplicando la fórmula de $T^2_{I1,I3}$

$$T^2_{I1,I3} = \frac{\displaystyle\sum_{i1=1}^{i1=m_1}\sum_{i3=1}^{i3=m_3} \frac{\left[n_{i1,i3}\right]^2}{n_{i1+}\; n_{+i3}} - 1}{\sqrt{(m_1 - 1)(m_3 - 1)}}$$

a la tabla de contingencia B_{13}

obtenemos el siguiente resultado:

$$T^2_{I1,I3} = \frac{\left(1+\frac{1}{3}+\frac{1}{3}+\frac{1}{3}\right) - 1}{\sqrt{(1)(3)}} = 0{,}5774$$

3º Cálculo de: $T^2_{I2,I3}$

Aplicando la fórmula de la $T^2_{I2,I3}$

$$T^2_{I2,I3} = \frac{\displaystyle\sum_{i1=1}^{i1=m_1}\sum_{i3=1}^{i3=m_3} \frac{\left[n_{i2,i3}\right]^2}{n_{i2+}\; n_{+i3}} - 1}{\sqrt{(m_2 - 1)(m_3 - 1)}}$$

a la tabla de contingencia B_{23}

obtenemos el siguiente resultado

$$T^2_{I2,I3} = \frac{\left(1+\frac{1}{2}+\frac{1}{2}+1\right) - 1}{\sqrt{(2)(3)}} = 0{,}8165$$

Representación gráfica de las variables cualitativas

A partir de los resultados del apartado anterior, estamos en condiciones de exponer la matriz **T cuadrado de Tschuprow**. Por tanto, podremos calcular los **valores propios** y los **vectores propios ortonormados,** mediante los métodos ya expuestos anteriormente.

Matriz T cuadrado de Tschuprow:

$$T^2 = \begin{pmatrix} 1{,}0000 & 0{,}2357 & 0{,}5774 \\ 0{,}2357 & 1{,}0000 & 0{,}8165 \\ 0{,}5774 & 0{,}8165 & 1{,}0000 \end{pmatrix}$$

Valores propios de la matriz T cuadrado de Tschuprow:

$$\lambda_1 = 2{,}1198536$$
$$\lambda_2 = 0{,}7792831$$
$$\lambda_2 = 0{,}1008632$$

Vectores propios ortonormados asociados a los valores propios de la matriz T cuadrado de Tschuprow.

$L_{11} = 0{,}4658595$ $L_{21} = 0{,}8242990$ $L_{31} = 0{,}3217236$

$L_{12} = 0{,}5832339$ $L_{22} = -0{,}5594750$ $L_{32} = 0{,}5889197$

$L_{13} = 0{,}6654421$ $L_{23} = -0{,}0867140$ $L_{33} = -0{,}7413960$

Representación gráfica de los puntos-variables en los círculos de correlaciones (1-2), (1-3) y (2-3).

PLANO (1-2)		
I1	**0,678278249**	**0,727666534**
I2	**0,849172054**	**-0,493887817**
I3	**0,968864867**	**-0,076548528**

PLANO (1-3)		
I1	**0,678278249**	**0,102176092**
I2	**0,849172054**	**0,187034814**
I3	**0,968864867**	**-0,235459712**

PLANO (2-3)		
I1	**0,727666534**	**0,102176092**
I2	**-0,493887817**	**0,187034814**
I3	**-0,076548528**	**-0,235459712**

En última instancia hemos de indicar que en 1979, Brigitte Escofier propuso un método para la representación de **variables cualitativas** en el **análisis de correspondencias múltiples** (48).

Análisis factorial de correspondencias múltiples(ACM) haciendo uso del AFC.

Primera situación:

El **número de modalidades** de las **variables cualitativas** es el mismo.

TABLA DE DATOS

$$T_{[BU]} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 3 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 2 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ \hdashline 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Mediante la aplicación de un AFC a la **tabla de datos,** podremos obtener una representación simultanea de las **modalidades** y de los **individuos** sobre los **planos factoriales** mas significativos, desde el punto de vista de la interpretación de la información. En el AFC las columnas representan las variables (I11, I12, I21, I22, I23, I31, I32, I33 e I34) y las filas representan los individuos (I11, I12, I21, I22, I23, I31, I32, I33, I34, I01, I02, I03 e I04)

El AFC de $T_{[BU]}$

conlleva a los siguientes resltados en cuanti a los **valores propios** y la **contribución de la inercia total.**

Valores propios	*Contribución de la inercia total*
0,7381	**52,4%**
0,4444	**31,6%**
0,2249	**16,0%**

Segunda situación:

El **número de modalidades** de las **variables cualitativas** no es el mismo.

Tabla de datos:

La matriz $T_{[BU]^o}$ *es la siguiente*

$$\left(\begin{array}{ccccccccc}
0{,}2171 & 0.0000 & 0{,}0000 & 0{,}1921 & 0.0000 & 0.1790 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.6512 & 0.1921 & 0.1921 & 0.1921 & 0.0000 & 0.1790 & 0.1790 & 0.1790 \\
0.0000 & 0.1921 & 0.7101 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.1584 & 0.0000 \\
0.1921 & 0.1921 & 0.0000 & 0.3401 & 0.0000 & 0.1584 & 0.0000 & 0.0000 & 0.1584 \\
0.0000 & 0.1921 & 0.0000 & 0.0000 & 0.1701 & 0.0000 & 0.1584 & 0.0000 & 0.0000 \\
0.1790 & 0.0000 & 0.0000 & 0.1584 & 0.0000 & 0.0000 & 0.3842 & 0.0000 & 0.0000 \\
0.0000 & 0{,}1790 & 0.0000 & 0.0000 & 0.1584 & 0.3842 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.1790 & 0.1584 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3842 \\
0.0000 & 0.1790 & 0.0000 & 0{,}1584 & 0.0000 & 0.0000 & 0.0000 & 0.3842 & 0.0000 \\
\hline
0.0000 & 0.4659 & 0.0000 & 0.0000 & 0.4124 & 0.0000 & 0.3842 & 0.0000 & 0.0000 \\
0.4649 & 0.0000 & 0.0000 & 0.4124 & 0.0000 & 0.3842 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.4649 & 0.0000 & 0{,}4124 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3842 \\
0.0000 & 0.4659 & 0.4124 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3842 & 0.0000
\end{array}\right)$$

Mediante la aplicación de un AFC a la **tabla de datos,** podremos obtener una representación simultanea de las **modalidades** y de los **individuos** sobre los **planos factoriales** mas

significativos, desde el punto de vista de la interpretación de la información. En el AFC las columnas representan las variables (I11*, I12*, I21*, I22*, I23*, I31*, I32*, I33* e I34*) y las filas representan los individuos (I11*, I12*, I21*, I22*, I23*, I31*, I32*, I32*, I33*, I34*, I1*, I2*, I3* e I4*)

El AFC de $T_{[BU]^o}$

conlleva a los siguientes resultados en cuanto a los **valores propios** y **contribución a la inercia total**

Valores propios	*Contribución a la inercia total*
0,6141	**31,0%**
0,3895	**19,6%**
0,3290	**16,6%**
0,3175	**16,0%**
0,2673	**13,5%**

Observación: estos resultados han sido obtenidos mediante la utilización del AFC debido a C.Dervin (49).Está contenido en el paquete de programas STATITCF.

ANEXO 1

La codificación lógica es insensible al arbitrio de los números de categoría

Introducción:

Esta aseveración está referenciada- pero no demostrada- en (50). Por tal motivo, incluimos su demostración y su aplicación en este anexo, por ser de gran importancia en el tratamiento estadístico de las encuestas de opinión.

En (50) la demostración, no está incluida, por ser un libro de la colección **Que sais-je?** cuya misión es, la de difundir –didácticamente– los contenidos de la misma para que, casi todas las personas –con un nivel mímino en matemática– puedan entenderlos.A título informativo, debo decirles que, en Espana tan sólo hay una revista que publica monografías de temas estadísticos. Nosotros ya hemos escrito dos libros en la Editorial La Muralla.

Demostración:

Para demostrar esta aseveración, hemos de hacer uso de la **matrices de permutación** y del **polinomio característico.**

1. Matrices de permutación y propiedades

Definición: una **matriz de permutación E** es tal que, cada fila y cada columna, contiene –únicamente– elementos nulos salvo un elemento que es igual a 1.

Propiedades:

$$E = E^T$$
$$E^{-1} = E^T$$

2. Operaciones con matrices de permutación

$$B = U^T\ U$$

$B\ C_{12}$: intercambia la colunma 1 por la 2 en la matriz B
$F_{12}\ B\ C_{12}$: intercambia la fila 1 por la 2 en la matriz $B\ C_{12}$

3. Polinomio característico

El **polinomio característico** de la **matriz de Burt** se define de la siguiente manera,

$$\det\left(B - \lambda I \right)$$

Demostrar que se verifica la siguiente igualdad:

$$\det\left(B - \lambda I \right) = \det\left(E^T\ B\,E - \lambda I \right)$$

$$\det\left(E^T\ B\,E - \lambda I \right) = \det\left(E\,B\,E - \lambda I \right) = \det\left(E^{-1}\ B\,E - \lambda\ E^{-1}\ E \right) =$$

$$= \det\left[E^{-1}\ (B - \lambda I)\,E \right] = \det\left(E^{-1} \right)\ \det\left(B - \lambda I \right)\ \det\left(E \right) =$$

$$= \det\left(B - \lambda I \right)$$

Por tanto,

$$\det\left(B - \lambda I \right) = \det\left(E^T\ B\,E - \lambda I \right)$$

A la vista de resultado, estamos en condiciones de afirmar que los **valores propios** de las matrices,

$$B \text{ y } E^T\ B\,E$$

son iguales.

A continuación, ejemplificaremos, con el caso de dos **situaciones** que, al ser distintas- en cuanto a la codificación se refiere- conducen a dos **matrices de Burt** las cuales, diagonalizándolas,llevarán a la conclusión de que, tienen los mismos **valores propios.**

SITUACIÓN 1

Variables cualitativas	*Modalidades*	*Codificación*
I1	I11 I12	1 2
I2	I21 I22 I23	1 2 3
I3	I31 I32 I33 I34	1 2 3 4

TABLA NUMÉRICA

Individuos	*I1*	*I2*	*I3*
1	2	3	2
2	1	2	1
3	2	2	4
4	2	1	3

TABLA DISYUNTIVA COMPLETA

Individuos	***I1***		***I2***			***I3***			
	I11	***I12***	***I21***	***I22***	***I23***	***I31***	***I32***	***I33***	***I34***
1	0	1	0	0	1	0	1	0	0
2	1	0	0	1	0	1	0	0	0
3	0	1	0	1	0	0	0	0	1
4	0	1	1	0	0	0	0	1	0

TABLA DE BURT

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 3 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 2 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

En donde, tanto las filas como las columnas representan las modalidades de las tres **variables cualitativas I1,I2 e I3.**

SEGUNDA SITUACIÓN

Variables cualitativas	*Modalidades*	*Codificación*
I1	I11 I12	2 1
I2	I21 I22 I23	3 2 1
I3	I31 I32 I33 I34	4 3 2 1

TABLA NUMÉRICA

Individuos	*I1*	*I2*	*I3*
1	1	1	3
2	2	2	4
3	1	2	1
4	1	3	2

TABLA DISYUNTIVA COMPLETA

Individuos	*I1*		*I2*			*I3*			
	I11	*I12*	*I21*	*I22*	*I23*	*I31*	*I32*	*I33*	*I34*
1	1	0	1	0	0	0	0	1	0
2	0	1	0	1	0	0	0	0	1
3	1	0	0	1	0	1	0	0	0
4	1	0	0	0	1	0	1	0	0

TABLA DE BURT

$$\begin{pmatrix} 3 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 2 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Conclusión:

Mientras que aplicando el

$$AFC\ de\ T_{[BU]}$$

con **tres ejes** retenemos el 100% de la **inercia total,** aplicando el

$$AFC\ de\ T_{[BU]^o}$$

con **cinco ejes** retenemos el 96,7% de la **inercia total.**

Se puede observar que, existe una gran diferencia entre los **valores propios** y su **contribución a la inercia total.**

El hecho de, la **ponderación de las matrices de variables indicadoras**, nos lleva a la retención de más **valores pripios** y, por consiguiente, tendremos que interpretar más **planos factoriales.**

Así pues, el método que proponemos, no es sólo más complicado que el usual –desde el punto de vista teórico– sino también, desde un punto de vista práctico. Sien embargo, sin lugar a dudas, las conclusiones, serán más exactas por las razones ya expuestas en este artículo.

AGRADECIMIENTOS

Este trabajo ha sido posible realizarlo gracias a una beca concedida por el Gobierno Francés.

Yves Escoufier (Ex-Presidente de la Universidad de Montpellier) me mostró el **Análisis de Datos,** según se practicaba en su Departmento de Biometría que dirigía en Montpellier. A él, le debo mi entusiasmo por este tema.He de añadir que, tuve ayuda de todos sus colegas. Gracias a gracias a José Luis Valencia Delfa –profesor Titular de **Análisis Estadístico Multidimensional** en la Escuela de Estadística de la Universidad Complutense de Madrid– pude seguir trabajando en este tema tan interesante sobre todo, para el establecimiento de tipologias concretas en el dominio de las ciencias experimentales. Agradezco a José Vicente Tarazona Lafarga, que es mi actual Director de Medio Ambiente en el INIA, por haberme estimulado y ayudado en la realización de este trabajo.

En última instancia, agradeco a José Luís de Miguel Arenal, Catedrático de Matemática I de la ESTIA de Madrid, no sólo los conocimientos de Análisis Estadístico Multidimensional que me transmitió en las asignaturas del Doctorado, sino también la ayuda constante que me estuvo proporcionando sobre dicho tema. Sin duda alguna él fue el que despertó en mí el interes sobre el Análisis Estadístico Multidimensional, parte integrante de la Matemática Aplicada.

BIBLIOGRAFÍA

(1) Wold,S.(1976).Pattern recognition by means of disjoint principal component models.Pattern Recognition, 8, p 127-139.

(2) Benzécri,J.P.(1977).Analyse discriminante et analyse factorielle.Les Cahiers de l'Analyse des Données, 4, p 369-406.

(3) Roux,M.(1985).Algorithmes de clasification. Masson.

(4) Escoufier, Y.(1970).Echantillonnage dans une population de variables réeles.Publication de l'Institut Statistique de l'Université de Paris, 19, fasc 4, p1-47.

(5) Escoufier,Y.(1973).Le traitement des variables vectorielles. Biometrics,vol 29, nº 4, pp 751-760.

(6) Escoufier,Y.;Robert,P.(1976).A unifying tood for linear multivariante methods:the RV coefficient.Applied Statistics. 25,(3),p 257-265.

(7) Escoufier,Y.(1979).Cours d'Analyse des Données.CRIG,av d'Occitanie 34075 Montpellier Cedex.

(8) Escoufier,Y;Robert,P.(1979).Choossing variables and metrics optimizing the RV coefficient.Academic Press INC-J.S.Rustagi.

(9) DO-CHI,Cl.B.(1979).Choix de variables en analyse des données.Thèse 3 ème cycle,USTL Montpellier.

(10) Gonzalez,P.L.(1981).Choix de variables, applications au choix de stations dans un réseau.RT. nº 8104,CRIG,av d'Occitanie 34075 Montpellier Cedex.

(11) Sabatier,R.(1981).Two examples of chossing variables and metrics with de RV coefficient. Rapport technique 8102, CRIG av.d'Occitanie 34075 Montpellier Cedex.

(12) Gonzalez,P.L.(1982).Analyse Statistique de Données psychosensorielles.Thèse présenté à l'Université des Sciences et Techniques du Languedoc pour obtenir le grade de Docteur de 3 ème cycle.Mathématiques pures et appliquées.

(13) Sabatier,R.(1983).Approximations d'un tableau de données. Applications à la reconstitution des paléoclimats.Thèse presenté à l'Université des Sciences et Techniques du Languedoc. pour obtenir le grade de Docteur de 3ème cycle.Mathématiques pures et appliquées.

(14) Díaz-Llanos Sainz-Calleja, Fco. J. (1985). Técnicas multidimensionales para el estudio de la evolución del sector agrario y afines a nivel provincial.Tesis presentada en la ETSIA de Madrid para obtener el grado de Doctor Ingeniero Agrónomo. Tesis dirigida por Jose Luís de Miguel Arenal- Catedrático de Matemática I.

(15) Díaz-Llanos Sainz-Calleja,Fco.J.;García Mouton,MªE.(1998). Una clasificación espacio-temporal, sin restricción de contigüidad geográfica, en regiones socioeconómicas. Estadística Española.Vol 40, Núm 143,pp 33-72.

(16) Rao,C.R.(1965). The use and interpretation of principal component analysis in applied research.Sankhya A,vol.26,p 329-58.

(17) Escoufier,Y.(1986).A propos du choix des variables en analyse des données.Metron,nº XLIV,p 31-47.

(18) Wilks,S.S.(1963).Multivariate Statistical outliers, Sankhya ,25,407-426.

(19) Cleroux, R.; Helbling, J. M.; Ranger, N. (1990). Détection d'ensembles de données aberantes en Analyse des Données Multivariées. Rev. Statistique Appliquée, XXXVIII(1),5-21.

(20) Crettaz de Roten,F.(1993).Données manquantes en statistique multivariée:une nouvelle méthode basée sur el coefficient RV.Thése n 1111.Ecole Polytechnique Fedérale de Lausanne.

(21) Hirschfeld,H.O.(1935).A connection between correlation and contingency. Proc. Camb. Philos.Soc.,31,pp. 520-524.

(22) Fisher,R.A.(1940).The precision of discriminant functions.Ann.Eugen (London) 10,pp. 422-429.

(23) Benzécri,J.P.(1976).Histoire et préhistoire de l'Analyse des Données. Les Cahiers de l'Analyse des Données 1, nº 1 à 4.Dunod.Paris.

(24) Benzécri,J-P.(1976).L'Analyse des Données.Tome 1:La Taxinomie;Tome 2:L'Analyse factorielle des correspondences. Dunod, Paris.

(25) Guttman,L.(1941).The quantification of a clas of attributes: a theory and method of a scale construction.In:The prediction of personal adjustment(Horst P.ed, p 251-264,SSCR New York.

(26) Burt,C.(1950).The factorial analysis of qualitative data.British J. Of. Statist. Psycol. 3,3, p 166-185.

(27) Hayashi,C.(1956).Theory and exemples of quantification.(II). Proc. of the Institute of Statist.Math.4(2),p 19-30.

(28) Lebart,L.(1975a).L'orientation du dépouillement de certaines enquêtes par l'analyse de correspondances multiples. Consommation,2,p 73-96.Dunod.

(29) Escofier,B.;Pagès,J.(1988)..Analyse factorielles simples et multiples.Objectifs, méthodes et interpretation.Dunod.

(30) Danbroise, E.; Escoufier, Y.; Massote, P. (1987). Application de l'Analyse des Données à l'elaboration de minisondages d'opinion. Revue de Statistique Appliquée, pp 9-23.

(31) Saporta,G.(1977).Une méthode et un programme d'analyse discrinimante sur variables qualitatives.In:Premières Journées Int.Analyse des Données et Informatique,INRIA,Rocquencourt.

(32) Kendal M.G.;Stuart A.(1958-1966).The advanced theory of statistics (3 tomes). Londres-Griffin

(33) Caillez,F.;Pages,J-P.(1976).Introduction à l'analyse des données. Smash.

(34) Foucart,T.(1984).Analyse factorielle de tableux multiples.Masson.

(35) Martín Pliego,Fco J.(1987).Curso práctico de Estadística económica.. Ed AC

(36) García Mouton,MªE.(1988).Nuevos métodos evaluativos multidimensionales para el tratamiento de las encuestas en el sector agrario.. Tesis presentada para optar al título de Doctor Ingeniero Agrónomo. Dirigida por Fco. J. Diáz-Llanos Sainz-Calleja.

(37) García Mouton,MªE.;Díaz-Llanos Sainz Calleja Fco. J.(1989). Efecto y utilidad de la métrica en el análisis de encuestas.Estadística Española.Vol 31, Núm 121, pp 253-280.

(38) Celeux,G.;Nakache,J-P.(1994). Analyse discriminante sur variables qualitatives. Polytechnique.

(39) 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 CEES (Centro Europeo de Estudios Superiores).ISBN .84-88881-22-3.

(40) Saporta,G.(1990).Probabilités.Analyse des données et statistique.Edition Technip.

(41) Daudin J.J.(1979).Coefficients de Tschuprow partiels et indépendance conditionnelle.Statistique et Analyse des Donées, 3,52-58.

(42) Householder,A.S.(1953).Principles of Numerical Analysis. Mc Graw-Hill, New York.

(43) Puy Huarte J.(1983).Cálculo numérico. Apuntes de Cátedra de Matemática III de la ETSI de Caminos,pp 320-342,399-402.

(44) Ciarlet P.G.(1985).Introduction à l'analyse numerique matricielle et á l'optimisation.Ed Masson, pp 90-94,123-131.

(45) Ralston A.(1978).Introducción al análisis numérico.Ed Limusa.Mexico.

(46) Diday, E.; Lemaire, J.; Pouget. J.; Testu, F. (1982). Éléments d'analyse de données. Dunod.

(47) Díaz-Llanos Sainz-Calleja, Fco. J. (2002). El análisis de datos en el cierre de ventas. Editorial La Muralla,SA. Espérides.

(48) Escofier,B.(1979).Une représentation des variables dans l'Analyse de Correspondandes Multiples.Revue de Statist. Appl ,27,pp 37-47.

(49) Dervin,C.(1990).Commment interpréter les résultats d'une analyse factorielle des correspondances.Institut Technique des Ceréales et des Fourrages.

(50) Cibois,Ph.(1983).L'Analyse factorielle. Que sais-je?.Presses Universitaires de France, pp 103-108.