

UN ANÁLISIS FACTORIAL DE LA ASOCIACIÓN DISIMÉTRICA ENTRE DOS VARIABLES CUALITATIVAS

RAFIK ABDESSELAM¹ – YVES SCHEKTMAN²

Resumen

Los “coeficientes de asociación relacionales” disimétricos son presentados. Las medidas de estos coeficientes se expresan en términos de inercias en el espacio de individuos provisto de un “producto escalar relacional”. Esta visión geométrica y mecánica de las asociaciones permite sintetizar y extender los métodos clásicos de análisis de datos, basados en la búsqueda de momentos principales de la nube de individuos, y permite proponer nuevos métodos. Para analizar la asociación disimétrica entre dos variables cualitativas, así como la determinación de los momentos principales y las representaciones gráficas, proponemos un análisis factorial aplicable a una familia de coeficientes de asociación disimétricos, incluyendo al tau de Goodman–Kruskal y sus extensiones: tau ponderado o equiponderado. Este análisis afina los resultados del análisis propuesto por D’Ambra y Lauro, y ofrece un campo de aplicación muy vasto. Además, es interesante notar que se encuentra el Análisis Factorial de Correspondencias de J.P. Benzécri, si se aplica este análisis al cuadrado medio de contingencia de Pearson. Se presenta un ejemplo sobre datos simulados.

Palabras–clave: Análisis factorial, distancia relacional, coeficiente de asociación simétrica y disimétrica, cuadrado medio de contingencia de Pearson, cociente de correlación, correlaciones canónicas, tau de Goodman–Kruskal, coeficiente de Stewart–Love.

Abstract

Asymmetrical “relational association coefficients” are described. Measurements of these coefficients are expressed as inertia in the individual–space with a “relational inner product”. This geometrical and mechanical point of view on associations analysis, leads to a synthesis, an extension, of classical data analysis methods, based on the research of principal axes of a configuration of points, and to new methods. We propose a factor analysis fitted to a family of asymmetrical association coefficients between two qualitative variables, including the Goodman–Kruskal tau and its weighted or equally weighted extensions. This analysis improves results proposed by D’Ambra and Lauro, and gives a wide scope of applications. Besides, it is interesting to note that Correspondence Factor Analysis is obtained by applying the proposed analysis to the symmetrical Pearson’s mean square contingency association coefficient. One example on simulated data is described.

¹LABORATOIRE LEMME–DIEM, UNIVERSITÉ PAUL SABATIER, TOULOUSE, FRANCIA

²MAISON DE LA RECHERCHE, UNIVERSITÉ TOULOUSE LE MIRAIL, FRANCIA

Key words: Factor analysis, relational distance, symmetrical and asymmetrical association coefficients, Pearson’s mean square contingency coefficient, correlation ratio, canonical correlations, Goodman–Kruskal tau, Stewart–Love coefficient.

1 Introducción – Notaciones

Se trata de analizar la asociación disimétrica “ x explica y ”, donde x y y son dos variables cualitativas que tienen respectivamente p y q modalidades. Para ello, se define un análisis factorial que describe la “forma de la asociación”, medida por un coeficiente escogido en una familia infinita de coeficientes de asociación disimétricos, que contienen al tau de Goodman–Kruskal [12], denotado τ . Este análisis afina los resultados del Análisis No Simétrico de Correspondencias propuesto por Lauro y D’Ambra [10, 14], y su campo de aplicación es más vasto. Además admite, como caso particular, el análisis de la asociación simétrica medida por el cuadrado medio de contingencia de Pearson, denotado ϕ^2 : por lo tanto, es equivalente al Análisis Factorial de Correspondencias (AFC) [4].

Sea $E_y = \mathbb{R}^q$ el subespacio de los individuos, asociado por dualidad a las variables indicatrices centradas de las modalidades de y , denotadas $\{y^k/k = 1, \dots, q\}$. El conjunto de los coeficientes de asociación disimétricos estudiados está en correspondencia biunívoca con el conjunto de los productos escalares de \mathbb{R}^q , es decir, el conjunto de matrices simétricas definidas positivas de orden q . La expresión general de estos coeficientes se da en la sección 2.3: se mide la inercia de una nube en el espacio de individuos $E = E_x \oplus E_y$ provisto de un producto escalar relacional [16], donde $E_x = \mathbb{R}^p$ es el subespacio de los individuos asociado por dualidad a las variables indicatrices centradas $\{x^j/j = 1, \dots, p\}$ de las modalidades de x . Presentaremos, en la sección 2, tres coeficientes: el tau de Goodman-Kruskal y sus derivados.

Se recuerda que un producto escalar relacional en E provee una traducción, en términos geométricos, de la estructura de las asociaciones observadas entre las variables $\{x^j\}$ y $\{y^k\}$. Los productos escalares relacionales son presentados suscintamente en la sección 2.3, ellos están en el origen de los métodos propuestos, sobre todo en [1, 2, 5, 8, 11, 13, 20].

La definición y las propiedades del análisis factorial propuesto son presentadas en la sección 3 y una aplicación sobre datos simulados es comentada en la sección 4.

Utilizaremos las siguientes notaciones:

- $X_{(n,p)}$ y $Y_{(n,q)}$ son las matrices de valores de las variables $\{x^j\}$ y $\{y^k\}$, donde n es el número de individuos, ponemos $V_x = X^tDX$, $V_y = Y^tDY$ y $V_{xy} = X^tDY$, donde $D = \frac{1}{n}I_n$ es la matriz diagonal de los pesos de los n individuos e I_n la matriz identidad de orden n ;
- M es la matriz del producto escalar de referencia en E y M_y [respectivamente M_x] es la del producto escalar del espacio, denotado E_y [resp. E_x], isomorfo al subespacio del mismo nombre, vía la inyección canónica denotada Iny [resp. Inx];
- $\mathcal{N}_y = \{y_i \in E_y/i = 1, \dots, n\}$ [resp. $\mathcal{N}_x = \{x_i \in E_x/i = 1, \dots, n\}$] es la nube de individuos asociada a Y [resp. X], está constituida por a lo sumo q (resp. p) puntos

distintos que representan también a las modalidades de y [resp. x];

- $g(y/x^j)$ es la q -upla $[(n_{jk}/n_{j\cdot}) - (n_{\cdot k}/n)/k = 1, \dots, q]$, donde $n_{j\cdot} = \sum_{k=1}^q n_{jk}$ y $n_{\cdot k} = \sum_{j=1}^p n_{jk}$, siendo n_{jk} el número de individuos que poseen las modalidades j de x y k de y ;
- $\mathcal{N}_g(y/x) = \{g(y/x^j)/j = 1, \dots, p\}$ es la nube de los p puntos $\{g(y/x^j)\}$ en \mathbb{R}^q ;
- D_y [resp. D_x] es la matriz diagonal (de pesos) definida por $[D_y]_{kk} = n_{\cdot k}/n$ para todo k (resp. $[D_x]_{jj} = n_{j\cdot}/n$ para todo j), donde de manera general $[A]_{kl}$ designa al elemento de la k -ésima fila y de la l -ésima columna de la matriz A ;
- D_{1/σ_y^2} es la matriz diagonal definida por $[D_{1/\sigma_y^2}]_{kk} = 1/\text{var}(y^k)$ para todo k , que también escribiremos $D_{1/\sigma_y^2} = \text{diag}[1/\text{var}(y^k)]$, donde $\text{var}(y^k)$ es la varianza de y^k ;
- $\chi_y^2 = D_y^{-1}$ [resp. $\chi_x^2 = D_x^{-1}$] es la matriz de la distancia del chi-cuadrado en \mathbb{R}^q [resp. \mathbb{R}^p].

Para simplificar las escrituras, se identificarán los vectores de los espacios E_x y E_y con su imagen por inyección canónica. Además, para no sobrecargar la notación, se usarán los mismos símbolos para las matrices y las aplicaciones que les corresponden en el esquema de dualidad [6]: así, el símbolo χ_y^2 servirá para denotar al producto escalar o el isomorfismo que le está asociado. Igualmente, el mismo símbolo designará un vector y la matriz de una columna de sus coordenadas en una base dada. Estos diferentes lenguajes serán utilizados indistintamente según su comodidad.

Se notará que $\mathcal{N}_g(y/x)$ es la nube de la p leyes condicionales centradas de y sabiendo x^j ($j = 1, \dots, p$), y que su centro de gravedad, respecto a la matriz de pesos D_x , se confunde con el origen de \mathbb{R}^q .

2 Expresiones de algunos coeficientes de asociación disimétricos

Tres enfoques son usados para dar las definiciones del τ y de sus derivados: al enfoque geométrico permite introducir naturalmente los derivados del τ , los enfoques algebraico y sobre todo el mecánico aportan las bases del método que proponemos.

2.1 Expresiones geométricas en el espacio de variables – Interpretación estadística

Sea $Q_x(y^k)$ la proyección ortogonal de $y^k \in \mathbb{R}^n$ sobre el subespacio generado por las $\{x^j/j = 1, \dots, p\}$. Como

$$\|Q_x(y^k)\|^2 = \sum_{j=1}^p \frac{n_{j\cdot}}{n} \left[\frac{n_{jk}}{n_{j\cdot}} - \frac{n_{\cdot k}}{n} \right]^2 \text{ y } \|y^k\|^2 = \frac{n_{\cdot k}(n - n_{\cdot k})}{n^2}, \quad (1)$$

entonces el tau de Goodman-Kruskal puede escribirse [19]:

$$\tau(x; y) = \frac{\sum_{k=1}^q \|Q_x(y^k)\|^2}{\sum_{l=1}^q \|y^l\|^2} \quad (2)$$

Entonces es razonable proponer las definiciones siguientes.

Definición 1 *Tau ponderado:*

$$\tau_p(x; y) = \frac{\sum_{k=1}^q n_{.k} \|Q_x(y^k)\|^2}{\sum_{l=1}^q n_{.l} \|y^l\|^2} \quad (3)$$

Tau equiponderado:

$$\tau_{ep}(x; y) = \frac{1}{q} \sum_{k=1}^q \frac{\|Q_x(y^k)\|^2}{\|y^k\|^2}. \quad (4)$$

Los términos utilizados, especialmente para el τ_{ep} , se justifican por las expresiones siguientes que se deducen inmediatamente de las expresiones (2), (3) y (4):

- $\tau(x; y) = \sum_{k=1}^q [\text{var}(y^k) / \sum_{l=1}^q \text{var}(y^l)] r^2[y^k; \{x^j\}]$
- $\tau_p(x; y) = \sum_{k=1}^q [n_{.k} \text{var}(y^k) / \sum_{l=1}^q n_{.l} \text{var}(y^l)] r^2[y^k; \{x^j\}]$
- $\tau_{ep}(x; y) = \frac{1}{q} \sum_{k=1}^q r^2[y^k; \{x^j\}]$

donde $r[y^k; \{x^j\}]$ es el coeficiente de correlación múltiple entre y^k y los $\{x^j / j = 1, \dots, p\}$.

Observación 1: si se desea usar un coeficiente de asociación disimétrica independiente de las varianzas de las $\{y^k\}$, entonces se podrá tomar τ_{ep} ; en el caso contrario se podrá tomar τ ó τ_p . La utilidad de τ_p se pone en evidencia en la sección 4.

2.2 Expresiones algebraicas

Para medir la asociación disimétrica entre x y y , es razonable proponer la fórmula siguiente:

$$F(x; y) = \sum_{j=1}^p \frac{n_{j.}}{n} \sum_{k=1}^q m_k(y) \left[\frac{n_{jk}}{n_{j.}} - \frac{n_{.k}}{n} \right]^2 \quad \text{con } m_k(y) \geq 0. \quad (5)$$

De la expresión (1) de $\|Q_x(y^k)\|^2$, se deduce que $F(x; y) = \sum_{k=1}^q m_k(y) \|Q_x(y^k)\|^2$; para establecer la propiedad siguiente, bastará entonces usar (2), (3), (4) y la expresión (1) de $\|y^k\|^2$.

Propiedad 1

- $F(x; y) = \tau(x; y)$ si $\forall k : m_k(y) = n^2 / \sum_{l=1}^q n_{.l}(n - n_{.l})$
- $F(x; y) = \tau_p(x; y)$ si $\forall k : m_k(y) = n^2 n_{.k} / \sum_{l=1}^q n_{.l}^2(n - n_{.l})$
- $F(x; y) = \tau_{ep}(x; y)$ si $\forall k : m_k(y) = n^2 / q n_{.k}(n - n_{.k})$.

Observación 2:

- a) En \mathbb{R}^q provisto de la distancia $M_y = \text{diag}[m_k(y)]$, respecto a D_x , los tres coeficientes anteriores tienen por valor la inercia, denotada I , de $\mathcal{N}_g(y/x)$ respecto al origen, o incluso respecto a su centro de gravedad. En efecto, (5) se escribe:

$$F(x; y) = \sum_{j=1}^p \frac{n_{j.}}{n} \|g(y/x^j)\|^2 = I[\mathcal{N}_g(y/x)].$$

- b) Si $\forall k : m_k(y) = n/n_{.k}$, es decir $M_y = \chi_y^2$, entonces $F(x; y) = \phi^2(x, y)$; además $F(x; y) = F(y; x)$ si $M_x = \chi_x^2$.
- c) Si $\forall k : n_{.k} = n/q$, es decir si las varianzas de las variables $\{y^k/k = 1, \dots, q\}$ son iguales, entonces $\tau(x; y) = \tau_{ep}(x; y) = \tau_p(x; y) = \phi^2(x, y)/(q - 1)$.

2.3 Expresiones mecánicas en el espacio de individuos provisto de una geometría relacional

La definición de “coeficientes de asociación relacionales” [1, 18] se basa en la noción de “producto escalar relacional”.

Ponemos $M_{xy} = \text{Inx}^t M \text{Iny}$, donde Inx^t es la transpuesta de Inx .

Se recuerda que en el espacio de individuos $E = E_x \oplus E_y$, el producto escalar M se llama relacional respecto a las variables $\{x^j\}$ y $\{y^k\}$, y se denota $\mathcal{R}[M_x, M_y]$, si y sólo si:

$$M_{xy} = M_x[(V_x M_x)^{1/2}]^\dagger V_{xy} M_y[(V_y M_y)^{1/2}]^\dagger \quad (6)$$

donde $[(V_x M_x)^{1/2}]^\dagger$ es la inversa generalizada de Moore-Penrose, ponderada por M_x , de $(V_x M_x)^{1/2}$.

La introducción de inversas generalizadas es una consecuencia de la singularidad de las matrices V_x y V_y , puesto que $\text{rang}(V_x) = p - 1$ y $\text{rang}(V_y) = q - 1$. Se muestra [22] que las propiedades de $\mathcal{R}[M_x, M_y]$ para matrices V_x y/o V_y singulares, son casi siempre idénticas a las establecidas para matrices regulares. Se efectúa, en anexo, una presentación sucinta de las inversas generalizadas ponderadas de Moore-Penrose y se muestran algunas propiedades útiles en Análisis de Datos.

Observación 3:

- a) Existen numerosas propiedades características de los productos escalares relacionales [8, 22]. Demos con fines ilustrativos una condición necesaria y suficiente de tipo geométrico [16]: sean $\{c_j(x)/j = 1, \dots, p\}$ [resp. $\{c_k(y)/k = 1, \dots, q\}$] los vectores axiales principales del triplete (N_x, M_x, D) [resp. (N_y, M_y, D)] y $\{C^j(x)/j = 1, \dots, p\}$ [resp. $\{C^k(y)/k = 1, \dots, q\}$] las componentes principales correspondientes, se tiene que $M = \mathcal{R}[M_x, M_y]$ es equivalente a:

$$\cos[c_j(x), c_k(y)] = \begin{cases} \cos[C^j(x), C^k(y)] & \text{si } \|C^j(x)\| \neq 0 \text{ y } \|C^k(y)\| \neq 0 \\ 0 & \text{sino.} \end{cases}$$

- b) A veces es útil [peso nulo de ciertos individuos, parejas de variables (x^j, y^k) muy correlacionadas] poder definir y usar semi-productos escalares relacionales: bajo ciertas condiciones [21, 24, 25] M_x, M_y y/o D pueden ser semi-productos escalares.

Se denota $I_x[\mathcal{N}_y]$ la inercia (respecto a su centro de gravedad) de la proyección ortogonal de \mathcal{N}_y sobre E_x , V_x^\dagger la inversa generalizada de Moore-Penrose de V_x (ponderada por M_x^{-1} y M_x) y P_x el operador de proyección ortogonal sobre E_x .

Lema 1 Si $M = \mathcal{R}[M_x, M_y]$ entonces $I_x[\mathcal{N}_y] = \sum_{k=1}^q \sum_{k'=1}^q [M_y]_{kk'} D[Q_x(y^k), Q_x(y^{k'})]$.

PRUEBA: Se sabe que

$$Q_x = XV_x^\dagger X^t D \text{ (cf. anexo) y } P_x = \text{Inx}, M_x^{-1} \text{Inx}^t M. \quad (7)$$

Partiendo de que $I_x[\mathcal{N}_y] = \text{traza}[P_x \text{Iny} V_y \text{Iny}^t M]$ y usando la igualdad (6), las descomposiciones espectrales de $[(V_x M_x)^{1/2}]^\dagger$ y $[(V_y M_y)^{1/2}]^\dagger$, y la relación $(V_x M_x)^\dagger = M_x^{-1} V_x^\dagger$ dadas en anexo, se tiene:

$$\begin{aligned} I_x[\mathcal{N}_y] &= \text{traza}[M_x^{-1} M_{xy} V_y M_{yx}] \\ &= \text{traza}[[(V_x M_x)^{1/2}]^\dagger V_{xy} M_y [(V_y M_y)^{1/2}]^\dagger V_y M_{yx}] \\ &= \text{traza}[V_{xy} M_y V_{yx} V_x^\dagger] \\ &= \text{traza}[M_y Y^t D Q_x Y] \\ &= \sum_{k=1}^q \sum_{k'=1}^q [M_y]_{kk'} D[Q_x(y^k), Q_x(y^{k'})]. \quad \blacksquare \end{aligned}$$

Observación 4: Es natural observar, en el lema 1, que $I_x[\mathcal{N}_y]$ no depende de la escogencia de M_x ; será a menudo útil escoger $M_x = \chi_x^2$, como será el caso de la definición 3 de la sección 3.2.

Definición 2 Sabiendo que $M = \mathcal{R}[M_x, M_y]$, se llama *Coficiente de Asociación Relacional Disimétrico* entre las variables $\{x^j\}$ y $\{y^k\}$, respecto a M_y , a la cantidad:

$$\text{CARD}[x; y] = I_x[\mathcal{N}_y] / I[\mathcal{N}_y].$$

Propiedad 2

$$\text{CARD}[x; y] = \begin{cases} \tau(x; y) & \text{si } M_y = I_q \\ \tau_p(x; y) & \text{si } M_y = D_y \\ \tau_{ep}(x; y) & \text{si } M_y = D_{1/\sigma_y^2}. \end{cases}$$

PRUEBA: Usando el lema 1 y las expresiones (2), (3) y (4), la propiedad proviene de los siguientes cálculos:

- Para $M_y = I_q$: $I_x[\mathcal{N}_y] = \sum_{k=1}^q D[Q_x(y^k), Q_x(y^k)] = \sum_{k=1}^q \|Q_x(y^k)\|^2$,

además $I[\mathcal{N}_y] = \text{traza}[V_y M_y] = \sum_{k=1}^q \text{var}(y^k) = \sum_{k=1}^q \|y^k\|^2$.

- Para $M_y = D_y$: $I_x[\mathcal{N}_y] = \sum_{k=1}^q \frac{n_{\cdot k}}{n} \|Q_x(y^k)\|^2$,

además $I[\mathcal{N}_y] = \sum_{k=1}^q \frac{n_{\cdot k}}{n} \|y^k\|^2$.

- Para $M_y = D_{1/\sigma_y^2}$: $I_x[\mathcal{N}_y] = \sum_{k=1}^q \|Q_x(y^k)\|^2 / \|y^k\|^2$,

además $I[\mathcal{N}_y] = q$. ■

Observación 5:

- a) Si las variables $\{x^j\}$ y $\{y^k\}$ son cuantitativas, entonces, para $M_y = I_q$, el valor del coeficiente $\text{CARD}[x; y]$ es igual [18] al del coeficiente de asociación disimétrica de Stewart-Love [23].
- b) Según el tipo y el número de variables, en los casos particulares en que $M_y = V_y^{-1}$ ó $M_y = \chi_y^2$, se observará que el numerador $I_x[\mathcal{N}_y]$ del coeficiente $\text{CARD}[x; y]$ es igual ([18], o ver prueba del lema 1) a la suma de los cuadrados de los coeficientes de correlación canónica, más aún, al cuadrado del coeficiente de Bravais-Pearson, al ϕ^2 de Pearson o al cociente de correlación generalizado.

El siguiente corolario se deduce de la demostración de la propiedad anterior y de la observación 4.

Corolario 1 *Sea M_x un producto escalar cualquiera en E_x . Si $M = \mathcal{R}[M_x, M_y]$ y*

- *si $M_y = (1/\sum_{l=1}^q \|y^l\|^2)I_q$, entonces $\tau(x; y) = I_x[\mathcal{N}_y]$.*

- *si $M_y = (1/\sum_{l=1}^q (n_{\cdot l}/n) \|y^l\|^2)D_y$, entonces $\tau_p(x; y) = I_x[\mathcal{N}_y]$.*

- si $M_y = (1/q)D_{1/\sigma_y^2}$, entonces $\tau_{ep}(x; y) = I_x[\mathcal{N}_y]$.

Si se denota $\text{diag}[a_l]$ una matriz diagonal cuyo l -ésimo elemento de la diagonal es a_l , se deduce del corolario 1, de las igualdades (1), de la propiedad 1 y de la observación 2(a), la propiedad que sigue.

Propiedad 3 Sea M_x un producto escalar cualquiera en E_x . Si $M = \mathcal{R}[M_x, M_y]$ y

- si $M_y = \text{diag}[n^2 / \sum_{l=1}^q n_{.l}(n - n_{.l})]$, entonces $\tau(x; y) = I_x[\mathcal{N}_y] = I[\mathcal{N}_g(y/x)]$.
- si $M_y = \text{diag}[n^2 n_{.k} / \sum_{l=1}^q n_{.l}^2 (n - n_{.l})]$, entonces $\tau_p(x; y) = I_x[\mathcal{N}_y] = I[\mathcal{N}_g(y/x)]$.
- si $M_y = \text{diag}[n^2 / q n_{.l}(n - n_{.l})]$, entonces $\tau_{ep}(x; y) = I_x[\mathcal{N}_y] = I[\mathcal{N}_g(y/x)]$.

3 Análisis factoriales de la asociación disimétrica entre dos variables cualitativas

Se usarán las siguientes notaciones:

- $\{e_j(x) \in E_x / j = 1, \dots, p\} \cup \{e_k(y) \in E_y / k = 1, \dots, q\}$ es la base canónica de $E = E_x \oplus E_y$ y $\{e_j^*(x) / j = 1, \dots, p\} \cup \{e_k^*(y) / k = 1, \dots, q\}$ su base dual;
- Pr_x (resp. Pr_y) es el proyector cartesiano de E sobre el espacio E_x (resp. E_y);
- si \underline{y}_i^k [resp. \underline{x}_i^j] es el valor de la indicatriz no centrada \underline{y}^k [resp. \underline{x}^j] para el individuo i , entonces ponemos $\underline{y}_i = \sum_{k=1}^q \underline{y}_i^k e_k(y)$ y $\underline{x}_i = \sum_{j=1}^p \underline{x}_i^j e_j(x)$;
- si y_i^k [resp. x_i^j] es el valor de la indicatriz centrada y^k [resp. x^j] para el individuo i , ponemos $y_i = \sum_{k=1}^q y_i^k e_k(y)$ y $x_i = \sum_{j=1}^p x_i^j e_j(x)$;
- si $\underline{X}_{(n,p)}$ es la matriz de elementos $[\underline{X}]_{ij} = \underline{x}_i^j$ para todo (i, j) , ponemos $V_{y\underline{x}} = Y^t D \underline{X}$;
- $g(\underline{x}) = \frac{1}{n} \sum_{i=1}^n \underline{x}_i = \sum_{j=1}^p \frac{n_{.j}}{n} e_j(x)$, $g(\underline{y}) = \frac{1}{n} \sum_{i=1}^n \underline{y}_i = \sum_{k=1}^q \frac{n_{.k}}{n} e_k(y)$;
- $g(x/y^k) = \sum_{j=1}^p [\frac{n_{jk}}{n_{.k}} - \frac{n_{j.}}{n}] e_j(x)$, es el centro de gravedad de los p puntos $\{e_j(x) - g(\underline{x}) / j = 1, \dots, p\}$, respecto al sistema de pesos $\{\frac{n_{jk}}{n_{.k}} / j = 1, \dots, p\}$;

- $\mathcal{N}_g(x/y) = \{g(x/y^k)/k = 1, \dots, q\}$;
- $g(y/x^j) = \sum_{k=1}^q [\frac{n_{jk}}{n_j} - \frac{n_{\cdot k}}{n}] e_k(y)$, es el centro de gravedad de los q puntos $\{e_k(y) - g(\underline{y})/k = 1, \dots, q\}$, respecto al sistema de pesos $\{\frac{n_{jk}}{n_j}/k = 1, \dots, q\}$;
- $\mathcal{N}_g(y/x) = \{g(y/x^j)/j = 1, \dots, p\}$.

Se observará que $\forall i \in \{1, \dots, n\}$, si $\underline{y}_i^k = 1$ entonces se tiene $\underline{y}_i = e_k(y)$ y $y_i = e_k(y) - g(\underline{y})$; así mismo $\forall i \in \{1, \dots, n\}$ $\underline{x}_i^j = 1$ implica $\underline{x}_i = e_j(x)$ y $x_i = e_j(x) - g(\underline{x})$.

Observación 6: Es inmediato mostrar que:

- $(\underline{X}^t D \underline{X})^{-1} = \chi_x^2$ y $V_{y\underline{x}} = V_{y\underline{y}}$.
- $g(\underline{x}) = \text{Inx} \underline{X}^t D \mathbf{1}_n$, donde $\mathbf{1}_n$ es el vector de coordenadas 1 en \mathbb{R}^n . (8)

- $g(y/x^j) = \sum_{k=1}^q [V_{y\underline{x}} \chi_x^2]_{kj} e_k(y) = \sum_{k=1}^q [V_{y\underline{y}} \chi_y^2]_{kj} e_k(y)$. (9)

3.1 Traducción del Análisis Factorial de Correspondencias en el modelo euclídeo relacional

Se sabe que el AFC es equivalente a los dos Análisis en Componentes Principales (ACP) siguientes:

$$\text{ACP del triplete } [\mathcal{N}_g(y/x); \chi_y^2; D_x] \quad (10)$$

y

$$\text{ACP del triplete } [\mathcal{N}_g(x/y); \chi_x^2; D_y]. \quad (11)$$

En (10) [resp. (11)] los vectores del espacio E_y [resp. E_x] han sido identificados con su imagen en el subespacio $E_y \subset E$ [resp. $E_x \subset E$], vía la inyección canónica Iny [resp. Inx].

Hacer un AFC se reduce entonces a efectuar una descomposición en momentos principales del valor de la asociación simétrica, entre las variables x y y , medida por el ϕ^2 de Pearson: el valor de este último es igual [cf. observaciones 2(a) y (b)] a

$$I[\mathcal{N}_g(y/x)] = I[\mathcal{N}_g(x/y)].$$

En $E = E_x \oplus E_y$ provisto del producto escalar relacional $\mathcal{R}[\chi_x^2, \chi_y^2]$, es decir en un Modelo Euclídeo Relacional (MER), recordemos [(8) y lemas 2 y 3 más abajo] que los momentos principales no nulos y las representaciones de los individuos (cada individuo está nombrado por el nombre de la modalidad correspondiente a la cualidad que él posee) de los dos ACP (10) y (11), son respectivamente idénticos (identificando los vectores de los espacios E_x, E_y con su imagen por la inyección canónica) a los de los dos ACP siguientes:

$$\text{ACP del triplete } [\{P_y(x_i)/i = 1, \dots, n\}; \mathcal{R}[\chi_x^2, \chi_y^2]; D] \quad (12)$$

y

$$\text{ACP del triplete } [\{P_x(y_i)/i = 1, \dots, n\}; \mathcal{R}[\chi_x^2, \chi_y^2]; D] \quad (13)$$

Se dirá, por extensión, que el ACP (12) [resp. (13)] es equivalente al ACP (10) [resp. (11)].

En (10) y (11), las nubes de puntos $N_g(y/x)$ y $N_g(x/y)$ están respectivamente en espacios \mathbb{R}^q y \mathbb{R}^p distintos, mientras que las de (12) y (13) están respectivamente en los subespacios $E_y = \mathbb{R}^q$ y $E_x = \mathbb{R}^p$ de un mismo espacio euclídeo $E = E_x \oplus E_y$, relacional para los conjuntos de variables $\{x^j\}$ y $\{y^k\}$.

Establezcamos los lemas siguientes en el marco más general en que los resultados son independientes del tipo de variables $\{y^k\}$. Para ello, se denota V_y^- [resp. V_x^-] una inversa generalizada interna de rango completo de V_y [resp. V_x] y se recuerda (cf. anexo) que χ_x^2 es una inversa generalizada de rango completo de V_x .

Lema 2 Si $M = \mathcal{R}[\chi_x^2, V_y^-]$ entonces $P_y[g(\underline{x})] = 0$ y $[\underline{x}_i^j = 1 \Rightarrow P_y(x_i) = P_y[e_j(x)] = g(y/x^j)]$.

PRUEBA: En vista de que $M = \mathcal{R}[\chi_x^2, V_y^-]$ se tiene $M_{yx} = V_y^- V_{yx} \chi_x^2$ (cf. anexo), usando (7) para P_y , se deduce:

$$P_y \text{Inx} = \text{Iny} V_{yx} \chi_x^2 = \text{Iny} V_{y\underline{x}} \chi_x^2. \quad (14)$$

Sabiendo que $Q_{\underline{x}} = \underline{X} \chi_x^2 \underline{X}^t D$ es una expresión del operador de proyección ortogonal sobre $\text{Im} \underline{X} \subset \mathbb{R}^n$, se deduce de (8) y (14),

$$P_y[g(\underline{x})] = \text{Iny} Y^t D Q_{\underline{x}} \mathbf{1}_n = \text{Iny} Y^t D \mathbf{1}_n = 0$$

puesto que $\mathbf{1}_n \in \text{Im} \underline{X}$ y que las $\{y^k\}$ están centradas;

lo anterior implica que $P_y(x_i) = P_y[\underline{x}_i - g(\underline{x})] = P_y(\underline{x}_i)$.

Finalmente, usando (14) y luego (9), y suponiendo $\underline{x}_i^j = 1$, se tiene

$$\begin{aligned} P_y(\underline{x}_i) = P_y[e_j(x)] &= \sum_{k=1}^q \langle \text{Iny} V_{yx} \chi_x^2 \text{Prx} e_j(x), e_k^*(y) \rangle e_k(y) \\ &= \sum_{k=1}^q [V_{yx} \chi_x^2]_{kj} e_k(y) = g(y/x^j). \quad \blacksquare \end{aligned}$$

Cuando y es cualitativa y las $\{x^j\}$ son de cualquier tipo, también se demostraría:

Lema 3 Si $M = \mathcal{R}[V_x^-, \chi_y^2]$ entonces $P_x[g(\underline{y})] = 0$, y $[\underline{y}_i^k = 1 \Rightarrow P_x(y_i) = P_x[e_k(y)] = g(x/y^k)]$.

Observación 7: Para evaluar las escogencias efectuadas, se observará (demostración similar a la propuesta en [22]) que (bajo condiciones de normalidad de M , no restrictivas en la práctica) si $M_x = V_x^-$ y $M_y = V_y^-$, entonces $\sum_{i=1}^n \frac{1}{n} \|y_i - P_y(x_i)\|_M^2$ (resp. $\sum_{i=1}^n \frac{1}{n} \|x_i - P_x(y_i)\|_M^2$) es mínimo si y sólo si $M = \mathcal{R}[V_x^-, V_y^-]$.

Las representaciones gráficas simultáneas y baricéntricas [4] de los puntos modalidades $\{x^j\} \cup \{y^k\}$, donde $\{x^j\}$ son los centros de gravedad de los $\{y^k\}$, son por lo tanto obtenidas, en el MER, con $\mathcal{R}[\chi_x^2, \chi_y^2]$, proyectando ortogonalmente sobre los planos principales del ACP (13) la nube:

$$\{P_x[e_k(y)]/k = 1, \dots, q\} \cup \{P_x P_y[e_j(x)]/j = 1, \dots, p\}. \quad (15)$$

Estas representaciones están justificadas por la igualdad de los momentos principales no nulos de los ACP (10) y (11), pero también por las fórmulas coherentes de transición [4] entre las componentes principales. Estas representaciones encuentran en el MER justificaciones suplementarias en (15) y en la propiedad siguiente.

Propiedad 4 Si $\{(\lambda_r, a_r)\}$ y $\{(\lambda_r, b_r)\}$ son los momentos principales no nulos y los vectores axiales principales respectivamente de los ACP (12) y (13), se tiene $P_x(a_r) = \sqrt{\lambda_r} b_r$.

PRUEBA: En efecto, se sabe (fórmulas de transición) que para todo momento principal no nulo λ_r , se tiene:

$$b_r = \frac{1}{\sqrt{\lambda_r}} \text{Inx } V_{xy} \chi_y^2 \text{Pr}_y a_r$$

ahora bien, usando la fórmula (14) para P_x , se tiene: $P_x(a_r) = \text{Inx } V_{xy} \chi_y^2 \text{Pr}_y a_r = \sqrt{\lambda_r} b_r$.

■

Observación 8:

- a) En AFC se pueden proponer otras representaciones, simultáneas y baricéntricas, permutando los roles de x y y , es decir proyectando la nube $\{P_y[e_j(x)]/j = 1, \dots, p\} \cup \{P_y P_x[e_k(y)]/k = 1, \dots, q\}$ sobre los planos principales del ACP (12).
- b) La propiedad 4 es una consecuencia de un caso particular ($M_x = \chi_x^2$ y $M_y = \chi_y^2$) de una propiedad característica de los productos escalares relacionales. En efecto,
 - se sabe [6, 7] que: $a_r = \text{Iny } Y^t D A^r$ y $b_r = \text{Inx } X^t D B^r$, donde (A^r, B^r) son las parejas de variables canónicas de (F_y, F_x, D) , siendo F_x [resp. F_y] el subespacio del espacio de variables generado por los vectores $\{x^j/j = 1, \dots, p\}$ [resp. $\{y^k/k = 1, \dots, q\}$];
 - las dos condiciones útiles [22] que caracterizan $M = \mathcal{R}[M_x, M_y]$ son:
 - i) los coeficientes de correlación canónica no nulos de (E_x, E_y, M) y (F_x, F_y, D) son iguales,
 - ii) las parejas de variables canónicas de mismo rango r ($a'_r \in E_y, b'_r \in E_x$) y ($A^r \in F_y, B^r \in F_x$) se corresponden por las relaciones siguientes:

$$a'_r = \text{Iny}[(V_y M_y)^{1/2}]^\dagger Y^t D A^r = \text{Iny } Y^t D A^r \text{ para } M_y = \chi_y^2$$

$$b'_r = \text{Inx}[(V_x M_x)^{1/2}]^\dagger X^t D B^r = \text{Inx } X^t D B^r \text{ para } M_x = \chi_x^2.$$

Se deduce $a'_r = a_r$ y $b'_r = b_r$, ahora bien $P_x(a'_r) = \sqrt{\lambda_r} b'_r$, de donde viene el resultado.

3.2 Definición

La propiedad 3 y los recordatorios anteriores nos llevan a proponer la siguiente definición.

Definición 3 *Se dice que se hace un análisis factorial de la asociación disimétrica entre la variable cualitativa x y la variable cualitativa y , medida por un coeficiente Q , si y sólo si:*

a) *se efectúan los dos ACP siguientes:*

$$\text{ACP del triplete } [\mathcal{N}_g(y/x); M_y; D_x] \quad (16)$$

y

$$\text{ACP del triplete } [\{P_x(y_i)/i = 1, \dots, n\}; \mathcal{R}[\chi_x^2, M_y]; D], \quad (17)$$

b) *el producto escalar M_y es tal que los valores de las inercias de las nubes de individuos de (16) y (17) sean iguales al valor del coeficiente de asociación disimétrica Q .*

A priori, M_x podría ser cualquiera, sin embargo la escogencia $M_x = \chi_x^2$ simplifica los cálculos y permite encontrar resultados fundamentales en el caso particular en que $M_y = \chi_y^2$. Por lo tanto, en lo que sigue supondremos que $M_x = \chi_x^2$.

Existencia: de acuerdo con la propiedad 3, existen productos escalares M_y para los coeficientes τ, τ_p y τ_{ep} . Más generalmente, se deduce de 1) de la propiedad 7 enunciada más abajo, que si Q es un CARD entonces siempre existe un producto escalar M_y : en efecto, si el coeficiente CARD ha sido definido respecto a $M = \mathcal{R}[M_x, M'_y]$, entonces $M_y = M'_y / I[\mathcal{N}_y]$, donde $I[\mathcal{N}_y]$ es calculada respecto a M'_y .

Observación fundamental: para $M_y = \chi_y^2$, el ACP (17) es idéntico al ACP (13) y por lo tanto equivalente al ACP (11); en este caso muy particular, el análisis factorial propuesto es entonces equivalente al AFC: se analiza la asociación simétrica, entre las variables x y y , medida por el ϕ^2 .

Proponemos entonces llamar al análisis correspondiente a la definición 3, el “Análisis Factorial de Correspondencias Disimétricas” (AFCD).

Puede darse una definición equivalente reemplazando (17) por el ACP equivalente que sigue:

$$\text{ACP del triplete } [\{P_x[e_k(\underline{y}) - g(\underline{y})]/k = 1, \dots, q\}; \mathcal{R}[\chi_x^2, M_y]; D_y]. \quad (18)$$

3.3 Propiedades

Se deduce de (9) la propiedad siguiente.

Propiedad 5 *El ACP (16) es idéntico al*

$$\text{ACP del triplete } [\chi_x^2 V_{xy}; M_y; D_x]. \quad (19)$$

Propiedad 6 *El ACP (17) es idéntico al*

$$\text{ACP del triplete } [Y M_y [(V_y M_y)^{1/2}]^\dagger V_{yx}; \chi_x^2; D]. \quad (20)$$

PRUEBA: Dado $M_x = \chi_x^2$, usando (6) y (7), sabiendo (cf. anexo) que $[(V_x \chi_x^2)^{1/2}]^\dagger V_{xy} = V_{xy}$, y $y_i = \text{Iny} Y^t f_i^*$ [6], donde $\{f_i^*/i = 1, \dots, n\}$ es la base dual de la base canónica del espacio de variables (\mathbb{R}^n) del ACP (17), se tiene:

$$P_x(y_i) = \text{Inx} M_x^{-1} M_{xy} Y^t f_i^* = \text{Inx} V_{xy} M_y [(V_y M_y)^{1/2}]^\dagger Y^t f_i^*.$$

Se deduce: $\langle P_x(y_i), e_j^*(x) \rangle = [V_{xy} M_y [(V_y M_y)^{1/2}]^\dagger Y^t]_{ji}$.

Las coordenadas de los puntos $\{P_x(y_i)/i = 1, \dots, n\}$, respecto a $\{e_j(x)/j = 1, \dots, p\}$, son por lo tanto elementos de las filas de la matriz $Y M_y [(V_y M_y)^{1/2}]^\dagger V_{yx}$, en vista de que $[(V_y M_y)^{1/2}]^\dagger$ es M_y -simétrica (cf. anexo). ■

Sean

- $\{\gamma_r(x)\}$ los momentos principales no nulos del ACP (19) ó (16), y $\{u_r\}$ los vectores axiales principales correspondientes, y
- $\{\gamma_r(y)\}$ los momentos principales no nulos del ACP (20) [resp. (17)] y $\{v_r\}$ [resp. $\{\text{Inx } v_r\}$] los vectores axiales principales correspondientes.

Propiedad 7 *Sea M_y un producto escalar cualquiera en E_y , entonces $\forall r$ se tiene:*

$$1) \quad \gamma_r(x) = \gamma_r(y) = \gamma_r \quad (21)$$

$$2) \quad u_r = \frac{1}{\sqrt{\gamma_r}} A v_r \text{ donde } A = V_{yx} \chi_x^2$$

$$v_r = \frac{1}{\sqrt{\gamma_r}} B u_r \text{ donde } B = V_{xy} M_y \quad (22)$$

PRUEBA: En vista de que $\chi_x^2 D_x = I_p$, los momentos y los vectores axiales principales del ACP (19) son los elementos propios (normados) del operador $(\chi_x^2 V_{xy})^t D_x \chi_x^2 V_{xy} M_y = AB$. Usando la M -simetría de $[(V_y M_y)^{1/2}]^\dagger$, los momentos y los vectores axiales principales del ACP (20) son los elementos propios (normados) del operador:

$$V_{xy} [(V_y M_y)^{1/2}]^\dagger M_y V_y M_y [(V_y M_y)^{1/2}]^\dagger V_{yx} \chi_x^2 = V_{xy} M_y V_{yx} \chi_x^2 = BA.$$

Como AB y BA tienen los mismos valores propios no nulos, se deducen fácilmente los asertos 1) y 2), teniendo en cuenta que u_r [resp. v_r] es normado por M_y [resp. χ_x^2]. ■

Observación 9: Al examinar las demostraciones de las propiedades 6 y 7, se constata que los resultados establecidos son válidos para cualquier tipo de variables $\{y^k\}$.

La siguiente propiedad es usada para poner en evidencia en la propiedad 9, unas fórmulas de transición, entre las componentes principales de (16) y (18), idénticas, para $M_y = \chi_y^2$, a las establecidas en AFC.

Propiedad 8 *El ACP (18) es equivalente al:*

$$\text{ACP del triplete } [\chi_y^2(V_y M_y)^{1/2} V_{yx}; \chi_x^2; D_y] \quad (23)$$

PRUEBA: Denotando $\tilde{y}_k = e_k(y) - g(y) = \text{Iny} \tilde{Y}^t f_k^*$, donde $\{f_k^*/k = 1, \dots, q\}$ es la base canónica del espacio de variables (\mathbb{R}^q) del ACP (18), se muestra, como en la propiedad 6, que las coordenadas de los puntos $\{P_x(\tilde{y}_k)/k = 1, \dots, q\}$, respecto a $\{e_j(x)/j = 1, \dots, p\}$, son los elementos de las filas de la matriz $\tilde{Y} M_y [(V_y M_y)^{1/2}]^\dagger V_{yx}$. Ahora bien, en vista de que:

- $\underline{y}_i^k = 1 \Rightarrow y_i = \tilde{y}_k$, se tiene $V_y = Y^t D Y = \tilde{Y}^t D_y \tilde{Y}$,
- $\langle e_k(y) - g(y), e_l^*(y) \rangle = [I_q - \mathbb{1}_q \mathbb{1}_q^t D_y]_{kl}$, se tiene $\tilde{Y} = I_q - \mathbb{1}_q \mathbb{1}_q^t D_y$, donde $\mathbb{1}_q$ es la matriz con una columna cuyos q elementos son iguales a 1 e I_q la matriz identidad de orden q ,

se deduce $V_y = [I_q - D_y \mathbb{1}_q \mathbb{1}_q^t] D_y \tilde{Y} = D_y \tilde{Y}$ pues las columnas de \tilde{Y} están centradas respecto a D_y .

Luego $\tilde{Y} = \chi_y^2 V_y$ y finalmente $\tilde{Y} M_y [(V_y M_y)^{1/2}]^\dagger V_{yx} = \chi_y^2 (V_y M_y)^{1/2} V_{yx}$. ■

Se denota $\{U^r\}$ [resp. $\{V^r\}$] las componentes principales del ACP (19) [resp. 23], correspondientes a los vectores axiales principales $\{u_r\}$ [resp. $\{v_r\}$]. Se recuerda que el ACP (23) es equivalente al ACP (18), el cual también es equivalente al ACP (17).

Propiedad 9 *Sea M_y un producto escalar cualquiera en E_y , entonces $\forall r$ se tiene:*

$$U^r = \frac{1}{\sqrt{\gamma_r}} \chi_x^2 V_{xy} [(V_y M_y)^{1/2}]^t V^r \quad (24)$$

$$V^r = \frac{1}{\sqrt{\gamma_r}} \chi_y^2 (V_y M_y)^{1/2} V_{yx} U^r \quad (25)$$

PRUEBA: Se tiene [6]:

$$U^r = \chi_x^2 V_{xy} M_y u_r \quad (26)$$

$$V^r = \chi_y^2 (V_y M_y)^{1/2} V_{yx} \chi_x^2 v_r. \quad (27)$$

Al reemplazar la expresión de v_r en (22) dentro de (27) y usando (26), se obtiene (25). Así mismo, al reemplazar la expresión de u_r en (21) dentro de (26), y usando (27), se tiene:

$$\begin{aligned} U^r &= \frac{1}{\sqrt{\gamma_r}} \chi_x^2 V_{xy} M_y V_{yx} \chi_x^2 v_r \\ &= \frac{1}{\sqrt{\gamma_r}} \chi_x^2 V_{xy} M_y [(V_y M_y)^{1/2}]^\dagger D_y V^r \\ &= \frac{1}{\sqrt{\gamma_r}} \chi_x^2 V_{xy} [(V_y M_y)^{1/2}]^t M_y (V_y M_y)^\dagger D_y V^r \\ &= \frac{1}{\sqrt{\gamma_r}} \chi_x^2 V_{xy} [(V_y M_y)^{1/2}]^t V_y^\dagger D_y V^r \text{ pues } (V_y M_y)^\dagger = M_y^{-1} V_y^\dagger \\ &= \frac{1}{\sqrt{\gamma_r}} \chi_x^2 V_{xy} [(V_y M_y)^{1/2}]^t V_y^\dagger V_y V^r \end{aligned}$$

ya que $V_y = Y^t D Y = \underline{Y}^t D Y = D_y - \underline{Y}^t D \mathbf{1}_n \mathbf{1}_q^t D_y$, y $\mathbf{1}_q^t D_y V^r = 0$ puesto que V^r está centrado;

así finalmente $U^r = \frac{1}{\sqrt{\gamma_r}} \chi_x^2 V_{xy} [(V_y M_y)^{1/2}]^t V^r$.

En efecto, $[(V_y M_y)^{1/2}]^t V_y^\dagger V_y = [V_y V_y^\dagger (V_y M_y)^{1/2}]^t = [(V_y M_y)^{1/2}]^t$ ya que (cf. anexo) $V_y V_y^\dagger$ es un operador de proyección sobre $\text{Im } V_y = \text{Im } Y^t = \text{Im } (V_y M_y)^{1/2}$ y $(V_y^\dagger)^t = V_y^\dagger$. ■

Observación 10: Si $M_y = \chi_y^2$, entonces (21), (22), (24) y (25) son las fórmulas de transición del AFC.

En AFCD, las representaciones simultáneas y baricéntricas, donde los $\{x^j\}$ son los centros de gravedad de los $\{y^k\}$ respecto al sistema de pesos $\{\frac{n_{jk}}{n_{j\cdot}}/k = 1, \dots, q\}$, se obtienen al proyectar ortogonalmente la nube $\{P_x[e_k(y) - g(y)]/k = 1, \dots, q\} \cup \{P_x[g(y/x^j)]/j = 1, \dots, p\}$ sobre los planos principales del ACP (17). Contrariamente al AFC, la disimetría de los papeles que juegan x y y no permite proponer aquí, para el coeficiente $Q(x; y)$, las representaciones obtenidas al permutar los roles de x y y (cf. observación 8): en efecto, estas representaciones serían las del AFCD para el coeficiente $Q(y; x)$.

En AFCD, los baricentros siempre son los representantes de las modalidades explicativas.

3.4 Ligamen con el Análisis No Simétrico de Correspondencias (ANSC)

En [10, 14], N. Lauro y L.D'Ambra definen el ANSC entre x y y por los dos ACP siguientes:

$$\text{ACP del triplete } [\{(\frac{n_{jk}}{n_{j\cdot}} - \frac{n_{\cdot k}}{n}; k = 1, \dots, q)/j = 1, \dots, p\}; M_y = I_q; D_x] \quad (28)$$

$$\text{ACP del triplete } [\{(\frac{n_{jk}}{n_{j\cdot}} - \frac{n_{\cdot k}}{n}; j = 1, \dots, p)/k = 1, \dots, q\}; M_x = D_x; I_q]. \quad (29)$$

Se tiene (propiedad 1):

$$[\sum_{l=1}^q \frac{n_{\cdot l}(n - n_{\cdot l})}{n^2}] \tau(x; y) = \sum_{j=1}^p \frac{n_{j\cdot}}{n} \sum_{k=1}^q \left[\frac{n_{jk}}{n_{j\cdot}} - \frac{n_{\cdot k}}{n} \right]^2 \quad (30)$$

$$= \sum_{k=1}^q \sum_{j=1}^p \frac{n_{j\cdot}}{n} \left[\frac{n_{jk}}{n_{j\cdot}} - \frac{n_{\cdot k}}{n} \right]^2. \quad (31)$$

Se observa que:

- (31) se obtiene a partir de (30) al permutar los símbolos de sumatoria.
- (30) es la expresión de la inercia de la nube del triplete (28), y (31) la de la nube del triplete (29), lo que, teniendo en cuenta el problema planteado, es coherente con las escogencias de los ACP (28) y (29).

Se constata que en el caso particular en que $M_y = I_q$,

a) El ACP (28) es idéntico al ACP (16) de la definición 3, para (cf. propiedad 3):

$$Q(x; y) = \left[\sum_{l=1}^q \frac{n_{.l}(n - n_{.l})}{n^2} \right] \tau(x; y).$$

b) El ACP (29) no es equivalente al ACP (17) ó (18): en efecto, se deduce de (9) que los momentos principales y los vectores axiales principales del ACP (29) son los elementos propios normados del operador $\chi_x^2 V_{xy} I_q V_{yx} \chi_x^2 D_x = \chi_x^2 V_{xy} V_{yx}$, mientras que el operador correspondiente del ACP (20) ó (23) es, para $M_y = I_q$, $V_{xy} V_{yx} \chi_x^2$ (cf. prueba de la propiedad 7).

Denotando (γ'_r, v'_r, V^r) los momentos principales no nulos, los vectores axiales principales y las componentes principales correspondientes del ACP (29), se tiene:

$$\forall r \quad \gamma'_r = \gamma_r, \quad v'_r = \chi_x^2 v_r$$

es más, teniendo en cuenta (27), se tiene

$$V^r = V_{yx} v'_r = V_{yx} \chi_x^2 v_r = [(V_y M_y)^{1/2}]^\dagger D_y V^r = [V_y^{1/2}]^\dagger D_y V^r \text{ para } M_y = I_q.$$

En conclusión:

- a) Para el análisis de la asociación disimétrica entre x y y , medida por el τ de Goodman-Kruskal, si bien es cierto que el ANSC y el AFCD parecen muy cercanos, especialmente en el caso muy particular en que $\chi_x^2 = pI_p$, de hecho los resultados gráficos del ACP (29) pueden ser muy diferentes de los del ACP (23), es decir de los del ACP (17) ó (18) para $M_y = I_q$ (cf. sección 4).
- b) En vista de que, en la definición 3, el producto escalar M_y es arbitrario, el AFCD ofrece un campo de aplicación más vasto que el del ANSC: así, para $M_y = \chi_y^2$ (caso simétrico) se identifica con el AFC, pero sobre todo, en el caso de una asociación disimétrica, permite efectuar un análisis tomando un coeficiente CARD “lo mejor adaptado posible” a la estructura observada de los datos (cf. sección 4).

4 Un ejemplo sobre datos simulados

4.1 Los datos

Se tiene $p = 5, q = 4$ y $n = 6391$; los datos se presentan en la tabla 1, donde $\{x^j/j = 1, \dots, 5\}$ [resp. $\{y^k/k = 1, \dots, 4\}$] representan las modalidades de la variable explicativa x [resp. a explicar y].

En las tablas 2, 3 y 4 se dan las cantidades que entran en las fórmulas de los coeficientes de asociación (cf. secciones 2.1 y 2.2), y que juegan por lo tanto papeles importantes en los análisis.

	x^1	x^2	x^3	x^4	x^5	$n_{.k}$
y^1	126	124	85	12	13	360
y^2	130	11	87	124	17	369
y^3	9	818	672	13	1352	2864
y^4	8	10	657	1065	1058	2798
$n_{j.}$	273	963	1501	1214	2440	6391

Tabla 1: Efectivos (n_{jk})

	x^1	x^2	x^3	x^4	x^5	$n_{.k}/n$
y^1	.4615	.1288	.0566	.0099	.0053	.0563
y^2	.4762	.0114	.0580	.1021	.0070	.0577
y^3	.0330	.8494	.4477	.0107	.5541	.4481
y^4	.0293	.0104	.4377	.8773	.4336	.4378

Tabla 2: Perfiles columna ($n_{jk}/n_{j.}$)

	x^1	x^2	x^3	x^4	x^5
y^1	.4052	.0725	.0003	-.0464	-.0510
y^2	.4185	-.0463	.0003	.0444	-.0507
y^3	-.4151	.4013	-.0004	-.4374	.1060
y^4	-.4085	-.4274	-.0001	.4395	-.0042

Tabla 3: Perfiles columna centrados ($n_{jk}/n_{j.} - (n_{.k}/n)$)

y^1	y^2	y^3	y^4
0.0531	0.0544	0.2473	0.2461

Tabla 4: Varianzas de las $\{y^k\}$

4.2 Algunos resultados numéricos

Se efectuaron cinco análisis:

- tres AFCD, respectivamente para los coeficientes τ_p , τ y τ_{ep} , denotados AFCD(τ_p), AFCD(τ) y AFCD(τ_{ep}), es decir respectivamente para:

$$- M_y = \text{diag}[n^2 n_{.k} / \sum_{l=1}^q n_{.l}^2 (n - n_{.l})],$$

$$- M_y = \text{diag}[n^2 / \sum_{l=1}^q n_{.l} (n - n_{.l})], \text{ y}$$

$$- M_y = \text{diag}[n^2 / q n_{.l} (n - n_{.l})];$$

- el AFC (o AFCD para el ϕ^2) y el ANSC con $M_y = I_q / [\sum_{l=1}^q \frac{n_{.l}(n - n_{.l})}{n^2}]$ (suma de los momentos principales igual a τ) con el fin de facilitar las comparaciones, sobre todo en la sección 4.3.2.

En la tabla 5, se dan algunos resultados numéricos obtenidos con los AFCD y el AFC: estos resultados corresponden a los gráficos de las figuras 1 y 2.

Análisis del tau ponderado $\tau_p(x; y) = 0.288$			
	1	2	Total
Valores propios	0.251	0.037	0.288
%	87.22	12.77	99.99
Contribuciones (%)			
x^1	0.01	87.72	11.21
x^2	40.69	0.44	35.55
x^3	0.00	0.00	0.00
x^4	57.42	0.03	50.09
x^5	1.88	11.81	3.15
y^1	0.05	44.14	5.68
y^2	0.14	44.44	5.80
y^3	49.74	5.36	44.08
y^4	50.07	6.06	44.44

Análisis del tau de Goodman-Kruskal $\tau(x; y) = 0.269$			
	1	2	Total
Valores propios	0.214	0.055	0.269
%	79.55	20.39	99.94
Contribuciones (%)			
x^1	0.00	87.71	17.89
x^2	40.93	0.41	32.66
x^3	0.00	0.00	0.00
x^4	57.27	0.02	45.58
x^5	1.80	11.86	3.87
y^1	0.59	44.59	9.59
y^2	0.79	44.00	9.63
y^3	49.07	5.47	40.15
y^4	49.55	5.94	40.63

Análisis del tau equiponderado $\tau_{ep}(x; y) = 0.231$			
	1	2	Total
Valores propios	0.137	0.094	0.231
%	59.33	40.51	99.84
Contribuciones (%)			
x^1	0.02	87.65	35.53
x^2	41.69	0.46	24.98
x^3	0.00	0.00	0.00
x^4	56.60	0.04	33.63
x^5	1.69	11.85	5.86
y^1	2.68	45.46	20.08
y^2	3.36	43.12	19.54
y^3	46.97	5.36	30.04
y^4	46.99	6.06	30.34

Análisis de correspondencias $\phi^2(x, y) = 0.645$			
	1	2	Total
Valores propios	0.327	0.318	0.645
%	50.59	49.20	99.79
Contribuciones (%)			
x^1	87.64	0.03	44.37
x^2	0.19	42.61	21.14
x^3	0.00	0.00	0.00
x^4	0.01	56.09	27.64
x^5	12.16	1.27	6.85
y^1	44.40	5.54	25.29
y^2	44.14	4.33	24.55
y^3	6.57	43.95	24.95
y^4	4.89	46.18	25.21

Tabla 5: Análisis Factorial de Correspondencias Disimétricas (τ_p, τ, τ_{ep}) y Simétricas (AFC: ϕ^2)

Los valores de las contribuciones de las modalidades, a los dos primeros momentos principales, son coherentes con las estadísticas de las tablas 2, 3 y 4: se notará especialmente que:

- Los dos grupos de parejas de modalidades bien asociadas $[(x^2; y^3), (x^4; y^4)]$ y $[(x^1; y^1), (x^1; y^2)]$ tienen papeles diferentes en los AFCD (τ_p, τ, τ_{ep}), y en el AFC (ϕ^2).
- Los valores de las contribuciones (Total) de las modalidades (y^3, y^4) a los coeficientes de asociación crecen partiendo de un mínimo observado en el AFC, pasando por el AFCD (τ_{ep}) y el AFCD (τ), para alcanzar un máximo observado en el AFCD (τ_p); así mismo, los valores de las contribuciones de estas mismas modalidades, al primer momento principal, en el AFCD (τ_{ep}) son inferiores a las observadas en el AFCD (τ), que son ellas mismas menores a las observadas en el AFCD (τ_p).

Con este juego de datos simulados, es muy claro que los resultados de los análisis de las asociaciones disimétricas medidas por los coeficientes $\{\tau_p, \tau, \tau_{ep}\}$ son diferentes de los del AFC.

En la tabla 6, se dan los mismos resultados, pero obtenidos con el ANSC: si se comparan estos resultados con los del AFCD (τ) (tabla 5), se observa: (i) que los valores propios y las contribuciones de las modalidades $\{x^j\}$ son (por definición) idénticas para los dos análisis, y (ii) que, al contrario, las contribuciones de las modalidades $\{y^k\}$ son diferentes, esencialmente las que conciernen el segundo momento principal. Estos resultados corresponden al gráfico de la figura 3.

Análisis del tau de Goodman-Kruskal $\tau(x; y) = 0.269$			
	1	2	Total
Valores propios	0.214	0.055	0.269
%	79.55	20.39	99.94
Contribuciones (%)			
x^1	0.00	87.71	17.89
x^2	40.93	0.41	32.66
x^3	0.00	0.00	0.00
x^4	57.27	0.02	45.58
x^5	1.80	11.86	3.87
y^1	0.63	25.28	5.68
y^2	0.74	24.72	5.66
y^3	49.78	24.56	44.61
y^4	48.85	25.44	44.05

Tabla 6: Análisis No Simétrico de Correspondencias

4.3 Algunos resultados gráficos

4.3.1 Gráficos

Un punto x^j [resp. y^k] de la figura 1(a) [resp. figura 1(b)] representa la proyección ortogonal del punto $g(y/x^j)$ [resp. $P_x[e_k(y) - g(\underline{y})]$] sobre el primer plano principal del ACP (16) [resp. (17) ó (18)] para $Q = \tau$.

Observamos inmediatamente, como hemos constatado sobre los resultados numéricos, que estos resultados gráficos son muy diferentes de los producidos por el AFC (cf. figura 2).

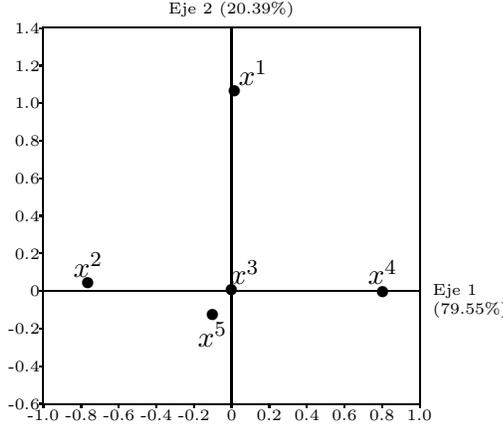
Sobre la figura 1(a) [resp. 1(b)]:

- entre más lejano esté un punto modalidad x^j [resp. y^k] del origen, más importante es su contribución al coeficiente de asociación disimétrica (aquí $Q = \tau$), por lo tanto más la modalidad “ x^j de x explica” [resp. “ y^k de y es explicada por”] la variable y [resp. x];
- entre más cercanos sean dos puntos, más las modalidades correspondientes de x [resp. y] explican y [resp. son explicadas por x] de manera similar.

El gráfico de la figura 1(a) [resp. figura 1(b)] es una descripción de los “potenciales a explicar” [resp. “potenciales para ser explicados”] de las modalidades de x [resp. y].

Así, como el porcentaje de inercia explicada por el primer plano principal es cercano al 100%,

(a) Análisis de la nube $\mathcal{N}_g(y/x)$
Tau de Goodman-Kruskal $\tau(x; y)$



(b) Análisis de la nube
 $\{P_x[e_k(y) - g(y)]; k = 1, \dots, 4\}$
Tau de Goodman-Kruskal $\tau(x; y)$

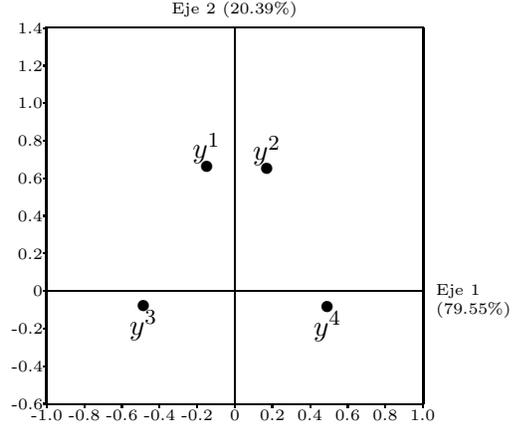


Figura 1: Análisis Factorial de Correspondencias Disimétricas para el τ de Goodman-Kruskal

- el potencial de previsión de x^3 es nulo, el de x^1 es el más elevado, y los de x^2 y x^4 son grandes y muy diferentes;
- los potenciales para ser explicados de y^1 y y^2 son más parecidos entre sí que los de y^3 y y^4 .

Estas afirmaciones son coherentes con la estructura de los perfiles columna centrados de la tabla 3.

En un AFCD, para representar gráficamente la asociación “ x explica a y ”, las representaciones simultáneas y baricéntricas (figura 2), definidas y justificadas al final de la sección 3.3, parecen ser, como en AFC, herramientas útiles.

Sobre la figura 3, se han representado los gráficos del ANSC, es decir los producidos por los ACP (28) y (29). Se observará que la nube de los $\{y^k\}$ tiene un “tamaño pequeño” respecto a la nube de los $\{x^j\}$: esto se debe al hecho que, por definición, (i) estas dos nubes tienen misma inercia, y (ii) que los puntos $\{y^k\}$ tienen todos masa igual a 1, mientras que los pesos de los $\{x^j\}$, más numerosos, son evidentemente más pequeños, respectivamente iguales a $\{0.04, 0.15, 0.24, 0.19, 0.38\}$.

Para facilitar la lectura, en ciertos gráficos, hemos indicado sobre las flechas, materializando la dirección de las asociaciones positivas, los valores de los perfiles columna centrados (tabla 3) correspondientes. Así mismo, para facilitar las comparaciones, las escalas sobre los ejes de las abscisas y de las ordenadas son las mismas.

Observación 11: Para representar gráficamente las asociaciones, sería también interesante estudiar las propiedades de los ACP siguientes:

1. ACP del triplete $[\{P_x[e_k(y) - g(y)]/k = 1, \dots, q\} \cup N_g(y/x)]; \mathcal{R}[\chi_x^2, M_y]; D_{y \cup x}]$

$$\text{con } [D_{y \cup x}]_l = \begin{cases} n_{.l}/2n & \text{para } l \leq q \\ n_{.l}/2n & \text{para } q + 1 \leq l \leq p + q. \end{cases}$$

2. ACP del triplete $[\{\{P_x[e_k(y) - g(\underline{y})]/k = 1, \dots, q\} \times N_g(y/x)\}; \mathcal{R}[\chi_x^2, M_y]; D_{q \times p}]$ donde $D_{q \times p}$ es una matriz diagonal de pesos de dimensión $q \times p$.

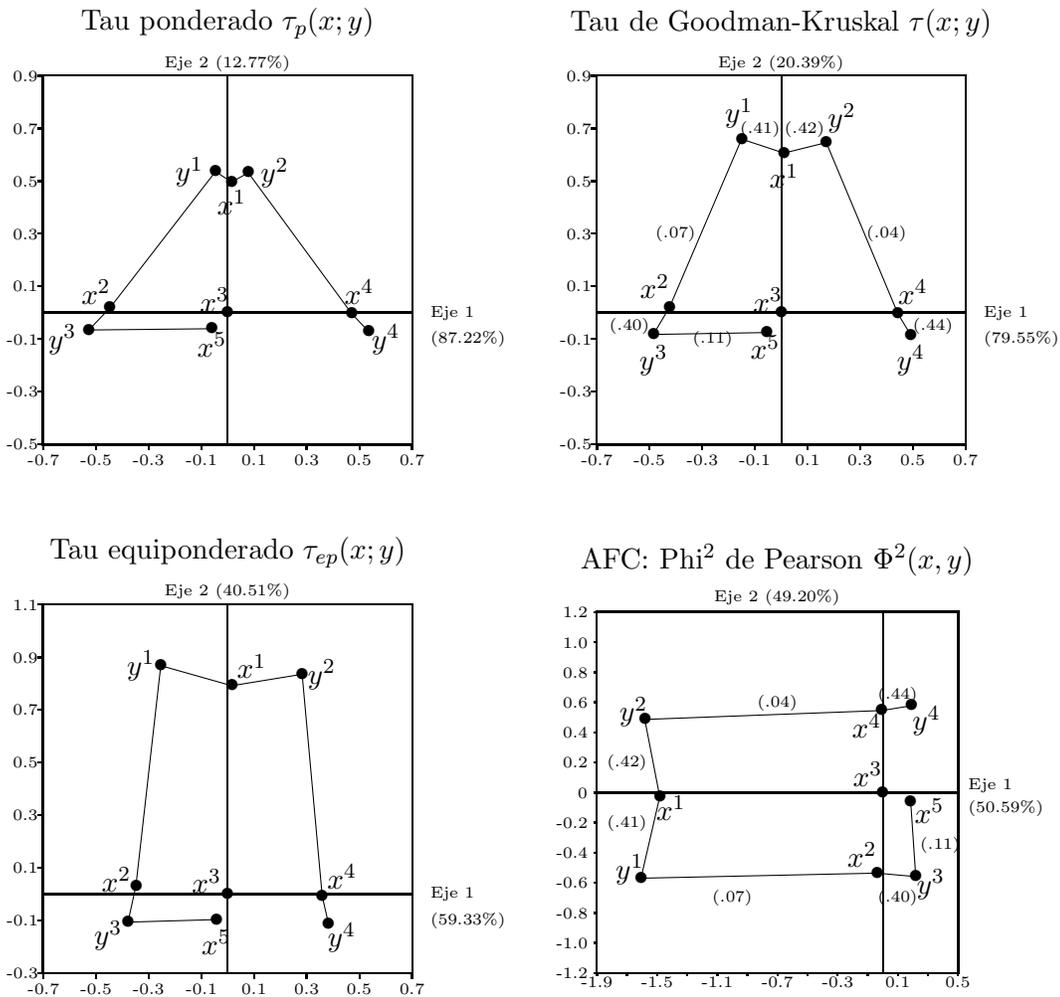


Figura 2: Análisis Factoriales de Correspondencias Disimétricas (τ_p, τ, τ_{ep}) y Simétricas (AFC: ϕ^2) (representaciones simultáneas y baricéntricas)

4.3.2 Algunos elementos para evaluar los análisis

Estas evaluaciones se basan en la comparación de los valores de las distancias, medidas sobre las representaciones gráficas (figuras 2 y 3) entre los puntos modalidades explicativas

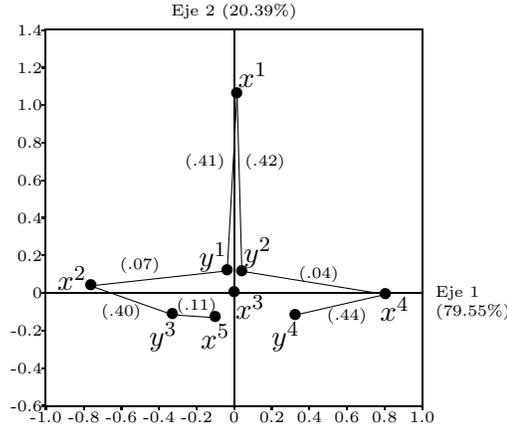
Tau de Goodman-Kruskal $\tau(x; y)$ 

Figura 3: Análisis No Simétrico de Correspondencias (representación simultánea)

$\{x^j\}$ y los puntos modalidades a explicar $\{y^k\}$ para un análisis dado, encontrándose los valores en la tabla 3.

En primer lugar, en vista de que los valores de la tercera columna de la tabla 3, correspondiente a x^3 , son iguales (nulos), se tiende a pensar que el gráfico del AFCD (τ_p) puede representar correctamente la asociación “ x explica a y ”, puesto que, para este análisis, las distancias de x^3 con los $\{y^k\}$, denotados $d(x^3, y^k)$, son aproximadamente iguales (tabla 7).

	AFCD			AFC	ANSC
	τ_p	τ	τ_{ep}		
$d(x^3, y^1)$	0.54	0.67	0.90	1.70	0.12
$d(x^3, y^2)$	0.54	0.67	0.88	1.65	0.12
$d(x^3, y^3)$	0.53	0.49	0.39	0.60	0.35
$d(x^3, y^4)$	0.54	0.50	0.40	0.61	0.34

Tabla 7: Distancias del punto modalidad explicativa x^3 a los puntos modalidades a explicar $\{y^k\}$.

Una segunda observación conduce a la misma conclusión.

Sea $\varepsilon = \{\varepsilon_{jk} = \frac{n_{jk}}{n_j} - \frac{n_{.k}}{n} / j = 1, \dots, p \text{ y } k = 1, \dots, q\}$,

$D = \{d(x^j, y^k) / j = 1, \dots, p \text{ y } k = 1, \dots, q\}$ el conjunto de las distancias [medidas sobre el plano principal (1,2)] entre las parejas (x^j, y^k) para un análisis dado y $C(\varepsilon, D)$ el índice de “concordancia” de rangos de Kendall entre las dos series de medidas ε y D : aquí el valor de $C(\varepsilon, D)$ es igual al número de productos negativos entre los $pq(pq - 1)/2$ productos de

la forma $(\varepsilon_{jk} - \varepsilon_{lm})[d(x^j, y^k) - d(x^l, y^m)]$, puesto que en una situación “ideal” se debería tener: $\varepsilon_{jk} > \varepsilon_{lm} \Rightarrow d(x^j, y^k) < d(x^l, y^m)$.

La lectura de la tabla 8 muestra que entre los valores del índice de concordancia $C(\varepsilon, D)$, calculado a partir de los gráficos de las figuras 2 y 3, la del AFCD (τ_p) es la más grande; se observará que el máximo de $C(\varepsilon, D)$ vale 190, puesto que $pq = 20$.

AFCD			AFC	ANSC
τ_p	τ	τ_{ep}		
164	161	140	140	114

Tabla 8: Valores del índice $C(\varepsilon, D)$ de concordancia de rangos de Kendall

Para evaluar visualmente, a partir de los valores de $C(\varepsilon, D)$, los resultados del AFCD (τ_p), del AFC y del ANSC, se han representado, sobre la figura 4, para cada uno de estos tres análisis, los gráficos de ε y de D en función de las parejas (x^j, y^k) ordenadas según el orden decreciente de los valores de los $\{\varepsilon_{jk}\}$ (cf. tabla 3).

Observación 12:

1. Se sugiere aquí una técnica, justificada y usada en otros lugares [1, 5, 11, 17, 24, 25], que ha permitido mejorar, en el sentido del criterio de concordancia [$C(\varepsilon, D) = 165$], los resultados obtenidos con el AFCD (τ_p): se ha escogido “el mejor” M_y [obtenido para $\alpha_2 = 14/15$] en una familia de distancias definida por:

$$M_y = (1 - \alpha_1)\chi_y^2 + \alpha_1 I_q \quad \text{y} \quad M_y = (1 - \alpha_2)I_q + \alpha_2 D_y$$

donde $\alpha_1, \alpha_2 \in \{0, 1/15, 2/15, \dots, 1\}$.

2. Hemos analizado con el AFCD (τ) los datos estudiados por F. Benzécri con el AFC en [3], y retomados por Lauro y D’Ambra [10] con el ANSC. No publicamos aquí estos resultados comparativos por falta de espacio; es más, estos últimos no presentan más que un interés medio: en efecto, aunque los momentos principales sean diferentes, los gráficos sobre el primer plano principal son muy parecidos.

5 Conclusión – Una apertura

La definición 3 y los resultados enunciados en el lema 2, las propiedades 6 y 7, siguen siendo válidos si las variables $\{y^k\}$ no son las indicatrices de una variable cualitativa. Los ACP (16) y (17) pueden por lo tanto ser usados para analizar la asociación disimétrica, medida por un coeficiente CARD, entre una variable cualitativa explicativa x y variables cuantitativas a explicar $\{y^k\}$. Este análisis engloba, en el caso particular en que $M_y = V_y^{-1}$ [cf. ACP (16)], al Análisis Factorial Discriminante (AFD) [6]: el coeficiente de asociación analizado es entonces simétrico pues el numerador del coeficiente CARD es igual al cociente de correlación generalizado [cf. observación 5 b)].

Se observará que las representaciones simultáneas y baricéntricas son las proyecciones ortogonales, sobre los planos principales del ACP (17), de la nube $\{P_x(y_i)/i = 1, \dots, n\} \cup$

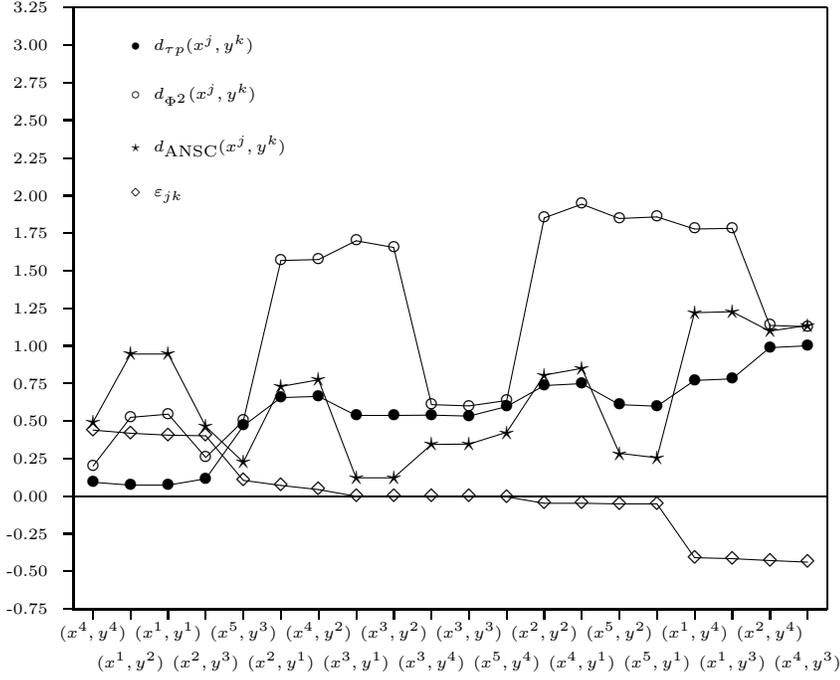


Figura 4: Gráfico de los $\varepsilon_{jk} = \frac{n_{jk}}{n_j} - \frac{n_{.k}}{n}$, y de las distancias $d(x^j, y^k)$ para el AFCD (τ_p), el AFC (ϕ^2) y el ANSC

$\{P_x[g(y/x^j)]/j = 1, \dots, p\}$, mientras que las del AFD son las proyecciones ortogonales, sobre los planos principales del ACP (16), de la nube $\{g(y/x^j) = P_y[e_j(x)]/j = 1, \dots, p\} \cup \{y_i/i = 1, \dots, n\}$. El procedimiento para analizar los resultados sería similar al presentado en las secciones 4.2 y 4.3.

Se notará, en fin, que el lema 2 y las propiedades 7 y 9 serían herramientas útiles para abordar el Análisis Multivariado de la Varianza (MANOVA) [5], o incluso para escribir ciertos algoritmos de clasificación [13].

Anexo: Algunos elementos sobre las inversas generalizadas ponderadas de Moore-Penrose[8,9,15]

A. Generalidades

Sean H y G dos espacios vectoriales provistos respectivamente de los productos escalares S y T , y sea A una aplicación de H en G .

Definición Se dice que A^\dagger es una Inversa Generalizada de Moore-Penrose (IGMP),

ponderada³ por la pareja (S, T) , de A si y sólo si:

$$AA^\dagger A = A \Leftrightarrow A^\dagger \text{ es la inversa generalizada interna (IGI) de } A \quad (32)$$

$$A^\dagger AA^\dagger = A^\dagger \Leftrightarrow A^\dagger \text{ es la inversa generalizada externa de } A \quad (33)$$

$$AA^\dagger [\text{resp. } A^\dagger A] \text{ es } T\text{-simétrica [resp. } S\text{-simétrica]} \quad (34)$$

Se deduce de (32) o (33) la idempotencia de los operadores AA^\dagger y $A^\dagger A$.

Es más, se tiene $\text{Im } AA^\dagger = \text{Im } A$ pues $\text{Im } A \supset \text{Im } AA^\dagger \supset \text{Im } AA^\dagger A = \text{Im } A$, y de la misma forma $\text{Im } A^\dagger A = \text{Im } A^\dagger$.

Se deducen las propiedades siguientes.

Propiedad I Las dos afirmaciones siguientes son equivalentes:

I.a) A^\dagger es la IGMP, ponderada por (S, T) , de A .

I.b) AA^\dagger es el operador de proyección T -ortogonal sobre $\text{Im } A$, $A^\dagger A$ es el operador de proyección S -ortogonal sobre $\text{Im } A^\dagger$.

Propiedad II Si A^- es una inversa generalizada interna de A , entonces AA^- es un proyector sobre $\text{Im } A$ paralelamente a $\ker AA^-$.

B. Algunas propiedades útiles en Análisis de Datos

Sean $X_{(n,p)}$ [resp. $\underline{X}_{(n,p)}$] la matriz de valores de las indicatrices centradas [resp. no centradas] de las p modalidades de una variable cualitativa observada sobre n individuos y $D = \frac{1}{n}I_n$ la matriz de pesos de los individuos, donde I_n es la matriz unidad de orden n . Se denota $\{(\lambda_j, u_j)/j = 1, \dots, p\}$ los valores propios y los vectores propios normados del operador VM del espacio de individuos \mathbb{R}^p , donde M es el isomorfismo asociado al producto escalar en \mathbb{R}^p y $V = X^t D X$. Se tiene $\text{rang } [V] = p - 1$, VM no es entonces invertible. Sin embargo, es útil tener una expresión de su IGMP, ponderada por M [es decir, por (M, M)], denotada $(VM)^\dagger$. Sea P_j el operador de proyección M -ortogonal sobre $\{\alpha u_j/\alpha \in \mathbb{R}\}$, en vista de que $\text{Im} [\sum_{\{j/\lambda_j \neq 0\}} P_j] = \text{Im } X^t$, se deduce inmediatamente de la

$$\text{propiedad I y de } VM = \sum_{\{j/\lambda_j \neq 0\}} \lambda_j P_j \text{ que } (VM)^\dagger = \sum_{\{j/\lambda_j \neq 0\}} \frac{1}{\lambda_j} P_j$$

De lo anterior se deduce que:

- Como P_j es M -simétrico, $(VM)^\dagger$ también lo es e $\text{Im } (VM)^\dagger = \text{Im } X^t$.
- $[(VM)^{1/2}]^\dagger = \sum_{\{j/\lambda_j \neq 0\}} \frac{1}{\sqrt{\lambda_j}} P_j$ es M -simétrica e $\text{Im } [(VM)^{1/2}]^\dagger = \text{Im } X^t$.
- $[(VM)^{1/2}]^\dagger (VM)^{1/2}$ es una expresión del operador de proyección M -ortogonal sobre $\text{Im } X^t$.

³También se dice inversa generalizada ortogonal

Sabiendo que P_j^t es el operador de proyección M^{-1} -ortogonal sobre $\{\alpha Mu_j/\alpha \in \mathbb{R}\}$, en vista de que $\text{Im} [M \sum_{\{j/\lambda_j \neq 0\}} \frac{1}{\lambda_j} P_j] = \text{Im} [\sum_{\{j/\lambda_j \neq 0\}} \frac{1}{\lambda_j} P_j^t M] = \text{Im} [\sum_{\{j/\lambda_j \neq 0\}} P_j^t]$ y notando que $V = \sum_{\{j/\lambda_j \neq 0\}} \lambda_j P_j M^{-1}$, se muestra, usando la propiedad I, que:

$V^\dagger = M \sum_{\{j/\lambda_j \neq 0\}} \frac{1}{\lambda_j} P_j$ es la IGMP, ponderada por (M^{-1}, M) , de V , con $\text{Im} VV^\dagger = \text{Im} X^t$.

Se deduce inmediatamente:

$$\text{a) } (V^\dagger)^t = \sum_{\{j/\lambda_j \neq 0\}} \frac{1}{\lambda_j} P_j^t M = \sum_{\{j/\lambda_j \neq 0\}} \frac{1}{\lambda_j} M P_j = V^\dagger.$$

$$\text{b) } (VM)^\dagger = M^{-1}V^\dagger.$$

c) $Q_x = XV^\dagger X^t D$ es una expresión del operador de proyección D -ortogonal sobre $\text{Im} X$: en efecto, la idempotencia de Q_x proviene del hecho que V^\dagger es una inversa generalizada externa, y la D -simetría proviene de la simetría de V^\dagger ; es más, $\text{Im} Q_x = \text{Im} X$ pues por una parte $\text{Im} Q_x \subset \text{Im} X$, y por otra parte $\text{traza}[Q_x] = \text{traza}[VV^\dagger] = \dim[\text{Im} VV^\dagger] = \dim[\text{Im} X^t]$.

Observación a: una inversa generalizada particular.

Partiendo de $\chi^2 = (\underline{X}^t D \underline{X})^{-1}$, si se denota $Q_{\mathbb{1}_n}$ el operador de proyección D -ortogonal sobre $\Delta_{\mathbb{1}_n}$, se tiene:

$$\begin{aligned} V\chi^2 V &= X^t D \underline{X} (\underline{X}^t D \underline{X})^{-1} \underline{X}^t D (I_n - Q_{\mathbb{1}_n}) \underline{X} \\ &= X^t D Q_{\underline{x}} \underline{X} - X^t D Q_{\underline{x}} Q_{\mathbb{1}_n} \underline{X} \\ &= V \end{aligned}$$

pues $Q_{\underline{x}} \underline{X} = \underline{X}$, $Q_{\underline{x}} Q_{\mathbb{1}_n} = Q_{\mathbb{1}_n}$ y $X^t D Q_{\mathbb{1}_n} \underline{X} = 0$.
Por lo tanto, χ^2 es una IGI de V .

De lo anterior se deduce que $V\chi^2$ es (cf. propiedad II) el proyector sobre $\text{Im} V = \text{Im} X^t$ paralelamente a $\ker V\chi^2$.

Observación b: una simplificación útil en $E = E_x \oplus E_y$.

Denotemos V_x^- [resp. V_y^-] una IGI de rango completo de V_x [resp. V_y]. En vista de que $V_x V_x^-$ es un proyector sobre $\text{Im} X^t$,

$$\text{si } M = \mathcal{R}[V_x^-, M_y] \text{ entonces } M_{xy} = V_x^- V_{xy} M_y [(V_y M_y)^{1/2}]^\dagger.$$

Se tiene evidentemente $M_{xy} = V_x^- V_{xy} V_y^-$ si $M = \mathcal{R}[V_x^-, V_y^-]$.

Referencias

- [1] Abdesselam, R. (1988) *Contribution à l'Analyse des Associations Dissymétriques*. Thèse de 3-ème cycle, Université Paul Sabatier, Toulouse, France.

- [2] Abdesselam, R., Schektman, Y. (1989) "Dissymmetrical association analysis between two qualitative variables", *Data Analysis, Learning Symbolic and Numeric Knowledge*, E. Diday *et al.* (eds.), INRIA - Nova Science Publishers, New York - Budapest: 39–46.
- [3] Benzécri, F. (1980) "Introduction à l'analyse des correspondances d'après un exemple de données médicales [INT.CORR.MED]", *Les Cahiers de l'Analyse des Données*, Vol. 5, No. 3: 283–310.
- [4] Benzécri, J.-P. (1982) *L'Analyse des Données: L'Analyse des Correspondances*, Tome 2. 4-ème Edition. Dunod, Paris.
- [5] Cadet, O.; Schektman, Y. (1989) "A method for analysing multidimensional experimental data", *Data Analysis, Learning Symbolic and Numeric Knowledge*, E. Diday *et al.* (eds.), INRIA - Nova Science Publishers, New York - Budapest: 87–94.
- [6] Cailliez, F.; Pagès, J.P. (1976) *Introduction à l'Analyse des Données*. SMASH-ASU-BURO, Paris.
- [7] Cazes, P. (1970) *Application de l'Analyse des Données au Traitement de Problèmes Géologiques*. Thèse de 3-ème cycle, Université Paris VI, France.
- [8] Croquette, A. (1980) *Quelques Résultats Synthétiques en Analyse des Données Multidimensionnelles: Optimalité et Métriques à Effets Relationnels*. Thèse de 3-ème cycle, Université Paul Sabatier, Toulouse, France.
- [9] Croquette, A. (1983) *Cours d'Analyse des Données*. Université Paul Sabatier, Toulouse, France.
- [10] D'Ambra, L.; Lauro, N. (1989) "Non symmetrical analysis of three-way contingency tables", *Multiway Data Analysis*, R. Coppi & S. Bolasco (eds.), North-Holland, Amsterdam: 301–315.
- [11] Fabre, C. (1986) *Contribution à la Protection des Méthodes Relationnelles*. Thèse de 3-ème cycle, Université Paul Sabatier, Toulouse, France.
- [12] Goodman, L.A.; Kruskal, W.H. (1954) "Measures of association for cross classifications", *J.A.S.A.* 49: 732–764.
- [13] Labrèche, S.; Schektman, Y.; Trejos, J.; Troupé, M. (1992) "Les distances relationnelles: deux applications récentes", *Congrès International sur Analyse en Distance, Distancia'92*, S. Joly & G. Le Calvé (eds.), Rennes: 369–372.
- [14] Lauro, N.; D'Ambra, L. (1983) "L'analyse non symétrique des correspondances", *Data Analysis and Informatics*, E. Diday *et al.* (eds.), North-Holland, Amsterdam: 433–446.
- [15] Nashed, M.Z., editor (1976) *Generalized Inverses and Applications*. Academic Press, New York.
- [16] Schektman, Y. (1978) *Contribution à la Mesure en Facteurs dans les Sciences Expérimentales et à la Mise en Œuvre des Calculs Statistiques*. Thèse de Doctorat d'État, Université Paul Sabatier, Toulouse, France.
- [17] Schektman, Y.; Fabre, C. (1984) "Un point de vue métrique sur la protection des modèles relationnels", *Résumés des Communications des Journées de Statistique*, Montpellier: 72.
- [18] Schektman, Y. (1988) "A general euclidean approach for measuring and describing associations between several sets of variables", *Recent Developments in Clustering and Data Analysis*, C. Hayashi *et al.* (eds.), Academic Press Inc., Tokyo: 37–48.
- [19] Schektman, Y. (1989) *Analyse et Traitement Informatique de Données*. Cours de D.E.A., Université Paul Sabatier, Toulouse.

- [20] Schektman, Y. (1989) “Inner products and association indices useful for analysing some multiway tables”, *Multiway Data Analysis*, R. Coppi & S. Bolasco (eds.), North-Holland, Amsterdam: 203–212.
- [21] Schektman, Y. (1991) “Eléments mathématiques de base pour la définition de semi-produits scalaires en analyse de données. Quelques applications”, Note interne, Laboratoire MLAD, Université Paul Sabatier, Toulouse.
- [22] Schektman, Y. (1994) “Propriétés des produits scalaires relationnels”, Note interne, Laboratoire Lemme-DIEM, Université Paul Sabatier, Toulouse (47 pp.)
- [23] Stewart, D.; Love, W. (1968) “A general canonical correlation index”, *Psychological Bull.*, Vol. 70: 160–163.
- [24] Trejos Zelaya, J. (1994) *Contribution à l'Acquisition Automatique de Connaissances à partir de Données Qualitatives*. Thèse, Université Paul Sabatier, Toulouse, France.
- [25] Troupé, M. (1994) *Contribution à la Régression Multiple Multidimensionnelle et à la Génération de Règles Incertaines*. Thèse, Université Paul Sabatier, Toulouse, France.

Agradecimientos

Agradecemos al Prof. P. Cazes por las observaciones y los consejos que nos brindó muy amablemente. También agradecemos al Prof. J. Trejos por la traducción del artículo.

N.B.: El contenido de este artículo será esencialmente publicado en la *Revue de Statistique Appliquée*, A.S.U., Institut Poincaré, 11 rue P. et M. Curie, Paris.