

# ALGORITMO E IMPLEMENTACIÓN DEL ANÁLISIS FACTORIAL DE CORRESPONDENCIAS

WILLIAM CASTILLO<sup>1</sup> – OLDEMAR RODRÍGUEZ<sup>1</sup>

---

## Resumen

En este artículo se presenta un algoritmo en pseudocódigo para el Análisis Factorial de Correspondencias. Se presentan también algunos comentarios sobre la implementación desarrollada en C++. Finalmente se ilustra el método mediante un ejemplo.

## Abstract

In this article, it is presented an algorithm in pseudo-code for the Factor Analysis of Correspondence. Some comments on the implementation on C++ are also developed. Finally, the method is illustrated by an example.

**Palabras clave:** Desarrollo de software, tabla de contingencia, perfiles, representación bidimensional, contribuciones.

## 1. Introducción

El Análisis Factorial de Correspondencias (AFC) es un método exploratorio concebido para estudiar la relación entre dos variables cualitativas a partir de la tabla de contingencia asociada. No obstante su uso se ha extendido a tablas de datos más generales, con tal que sus casillas sean no negativas y sumables por filas y columnas [4] y [5].

El AFC fue implementado como un módulo del sistema informático de análisis de datos PIMAD 2.1 [6], utilizando para ello el modelo matemático del Análisis en Componentes Principales (ACP) y los resultados específicos que se obtienen en el caso particular del AFC, como se expone en la sección siguiente.

En este artículo se presentan además los índices de calidad usuales (sección 3) tal como fueron implementados en el programa. El artículo se completa con una formulación algorítmica en pseudocódigo a partir de la cual se realizó la implementación computacional del método; luego se ofrecen algunos detalles sobre el software desarrollado y en la última sección se ilustra, a partir de un ejemplo, el uso de este software.

---

<sup>1</sup>PIMAD-CIMPA, ESCUELA DE MATEMÁTICA, UNIVERSIDAD DE COSTA RICA, 2060 SAN JOSÉ, COSTA RICA

## 2. El modelo matemático de base

Uno de los objetivos del AFC es explorar las relaciones entre las modalidades de dos variables cualitativas por medio de representaciones bidimensionales óptimas. Como se sabe, este objetivo se logra haciendo el ACP de las nubes de perfiles fila y columna, lo cual permite obtener tales representaciones. Sean  $X$  y  $Y$  dos variables cualitativas con  $n$  y  $p$  modalidades respectivamente,  $K_{n \times p}$  la tabla de contingencia asociada y  $S = \frac{1}{T}K$  la tabla de frecuencias; donde  $T = \sum_{ij} k_{ij}$ .

### 2.1. Perfiles y distancias

Sean  $f_i = \sum_{j=1}^p k_{ij}$ ,  $c_j = \sum_{i=1}^n k_{ij}$ . Los perfiles fila y columna son, respectivamente, los vectores  $pf_i = \left(\frac{k_{i1}}{f_i}, \dots, \frac{k_{ip}}{f_i}\right)$  y  $pc_j = \left(\frac{k_{1j}}{c_j}, \dots, \frac{k_{nj}}{c_j}\right)^t$ . A cada perfil  $pf_i$  ( resp.  $pc_j$ ) se le asigna el peso  $p_i = \frac{f_i}{T}$  ( resp.  $q_j = \frac{c_j}{T}$ ).

Las distancias Chi-cuadrado entre los perfiles fila y los perfiles columna son dadas, respectivamente por las matrices  $D_q^{-1}$  y  $D_p^{-1}$ ; donde  $D_p = \text{diag}(p_i)$  y  $D_q = \text{diag}(q_j)$ . Es decir:

$$d^2(pf_i, pf_l) = \sum_{j=1}^p \frac{T}{c_j} \left( \frac{k_{ij}}{f_i} - \frac{k_{lj}}{f_l} \right)^2$$

$$d^2(pc_j, pc_h) = \sum_{i=1}^n \frac{T}{f_i} \left( \frac{k_{ij}}{c_j} - \frac{k_{ih}}{c_h} \right)^2.$$

### 2.2. ACP de los perfiles

Las filas de las matrices  $D_p^{-1}S$  y  $D_q^{-1}S^t$  son respectivamente los perfiles fila y columna. Por lo tanto, identificando cada modalidad con su perfil y haciendo el ACP de los tripletes  $(D_p^{-1}S, D_q^{-1}, D_p)$  y  $(D_q^{-1}S^t, D_p^{-1}, D_q)$ , se obtienen las representaciones óptimas como proyecciones ortogonales de las modalidades (perfiles), sobre los espacios principales correspondientes ([1], [3] y [5]).

Se sabe que los elementos propios de estos dos ACP's se relacionan de modo tal que los valores propios son los mismos y los vectores propios de un ACP se calculan conociendo los del otro. Este resultado es la base teórica del algoritmo que se presenta en la sección 3 y es lo que permite una implementación computacional eficiente del AFC. Una expresión formal de tales resultados es la siguiente:

1. Los valores propios de los dos ACP's son iguales. Los vectores propios están relacionados así:  $u_i = \frac{1}{\sqrt{\lambda_i}} S^t D_p^{-1} v_i$  y  $v_i = \frac{1}{\sqrt{\lambda_i}} S D_q^{-1} u_i$  donde  $u_1, \dots, u_a$  y  $v_1, \dots, v_a$  son, respectivamente, los vectores propios ortonormados de los ACP 's de los tripletes  $(D_p^{-1}S, D_q^{-1}, D_p)$  y  $(D_q^{-1}S^t, D_p^{-1}, D_q)$ , asociados a los valores propios  $\lambda_1, \dots, \lambda_a$ , con  $a = \min\{n, p\}$ .
2. Con la notación anterior se tiene  $\lambda_1 = 1$ ,  $u_1 = G_f$ ,  $v_1 = G_c$  (salvo por el signo).

3. Fórmulas baricéntricas<sup>2</sup>: sea  $\text{coord}_{u_\alpha}(pf_i) = pf_i \cdot D_q^{-1} \cdot u_\alpha$  (resp.  $\text{coord}_{v_\alpha}(pc_j) = pc_j \cdot D_p^{-1} \cdot v_\alpha$ ) la coordenada del perfil  $pf_i$  (resp.  $pc_j$ ) sobre la recta determinada por el vector  $u_\alpha$ . (resp.  $v_\alpha$ ). Las fórmulas baricéntricas toman la forma,

$$a) \quad \text{coord}_{u_\alpha}(pf_i) = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{j=1}^p \frac{k_{ij}}{f_i} \text{coord}_{v_\alpha}(pc_j) \text{ y}$$

$$b) \quad \text{coord}_{v_\alpha}(pc_j) = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{i=1}^n \frac{k_{ij}}{c_j} \text{coord}_{u_\alpha}(pf_i).$$

4.  $\lambda \in ]0, 1]$  para todo valor propio no nulo del AFC. Este resultado se puede obtener directamente como una consecuencia de las fórmulas baricéntricas ([2] y [5]).
5. Sea  $\widehat{PS}$  la matriz  $D_p^{-1}S$  centrada por columna y  $\widehat{V}_f = \widehat{PS}^t D_p \widehat{PS}$ . Si  $Pr_{u_\alpha}$  es el operador de proyección  $D_q^{-1}$ -ortogonal sobre  $u_\alpha$  y  $Pr_{v_\alpha}$  el operador de proyección  $D_p^{-1}$ -ortogonal sobre  $v_\alpha$ , entonces<sup>3</sup>:

$$(a) \quad \widehat{V}_f D_q^{-1} u_i = \lambda_i u_i, \quad i = 2, \dots, a.$$

$$(b) \quad Pr_{u_\alpha}(pf_i - G_f) = Pr_{u_\alpha}(pf_i); \quad i = 1, \dots, n.$$

$$(c) \quad Pr_{v_\alpha}(pc_j - G_c) = Pr_{v_\alpha}(pc_j); \quad j = 1, \dots, p.$$

1. Simetrización<sup>4</sup>:  $V_f D_q^{-1} \sim D_q^{-\frac{1}{2}} (V_f D_q^{-1}) D_q^{\frac{1}{2}} = D_q^{-\frac{1}{2}} S^t D_p^{-1} S D_q^{-\frac{1}{2}} = Z$  que es simétrica. Por lo tanto  $Z$  y  $V_f D_q^{-1}$  tienen los mismos valores propios y la relación entre los vectores propios es  $v = D_q^{-\frac{1}{2}} u$ , donde  $u$  y  $v$  son vectores propios de  $V_f D_q^{-1}$  y  $Z$  respectivamente, asociados a  $\lambda$ .

### 3. Interpretación de un AFC

La interpretación de un Análisis Factorial de Correspondencias tiene que ver con la selección de ejes factoriales significativos, de puntos significativos y su representación plana; todo lo cual permite visualizar las proximidades y oposiciones entre perfiles. A continuación se presentan los índices y criterios utilizados como ayudas usuales en la etapa de análisis e interpretación de las “salidas” en un programa de AFC. La implementación en PIMAD permite hacer uso de estos índices y criterios.

<sup>2</sup>A nivel de la implementación computacional del AFC, estas fórmulas permiten hacer todos los cálculos a partir de uno solo de los dos ACP 's.

<sup>3</sup>Según este resultado el AFC se puede implementar sin hacer la operación de centraje con tal que no se tome en cuenta el primer valor propio.

<sup>4</sup>El algoritmo de diagonalización requiere que la matriz sea simétrica, por eso se diagonaliza  $Z$  y luego se hacen los ajustes necesarios en el cálculo de coordenadas, de acuerdo con este resultado.

### 3.1. La contribución absoluta

La contribución absoluta es un indicador del aporte inercial de una modalidad a un eje. Como se sabe, cada eje explica una parte de la inercia de la nube de perfiles, que es justamente la inercia de la nube proyectada sobre este eje. Así por ejemplo, para el eje  $\alpha$ -ésimo se tiene:

$$\text{Inercia proyectada} = \lambda_\alpha = \sum_{i=1}^n p_i (\text{coord}_{u_\alpha}(pf_i))^2 = \sum_{j=1}^p q_j (\text{coord}_{v_\alpha}(pc_j))^2.$$

En virtud de esta relación es natural definir la **contribución absoluta** del perfil  $pf_i$  al eje  $\alpha$ -ésimo, como

$$\text{ctr}_\alpha(i) = \frac{p_i (\text{coord}_{u_\alpha}(pf_i))^2}{\lambda_\alpha} = \frac{f_i (\text{coord}_{u_\alpha}(pf_i))^2}{T\lambda_\alpha}.$$

Cuanto mayor es  $\text{ctr}_\alpha(i)$ , más contribuye el perfil a la construcción del eje  $\alpha$ -ésimo. De la misma manera la contribución de un perfil columna  $pc_j$  al eje  $\alpha$ -ésimo es

$$\text{ctr}_\alpha(j) = \frac{q_j (\text{coord}_{v_\alpha}(pc_j))^2}{\lambda_\alpha} = \frac{c_j (\text{coord}_{v_\alpha}(pc_j))^2}{T\lambda_\alpha}$$

### 3.2. Contribución relativa

La **contribución relativa** se puede usar como un índice para evaluar la calidad de la representación de las modalidades en los ejes y planos principales. También puede ser usado para dar significado a un eje con el cual se relacionan ciertas modalidades. Como  $\|pf_i - G_f\|^2 = \sum_{\alpha=2}^a \|\text{Pr}_{u_\alpha}(pf_i)\|_{D_q^{-1}}^2$  entonces es natural definir la contribución relativa del eje  $\alpha$ -ésimo al perfil  $pf_i$ , como:

$$\text{corr}^2(\theta_{i\alpha}) = \frac{\|\text{Pr}_{u_\alpha}(pf_i)\|_{D_q^{-1}}^2}{\|(pf_i - G_f)\|_{D_q^{-1}}^2} = \frac{[\text{coord}_{u_\alpha}(pf_i)]^2}{\sum_{j=1}^p \frac{T}{c_j} \left( \frac{k_{ij}}{f_i} - \frac{c_j}{T} \right)^2}$$

el cual se interpreta geométricamente como el cuadrado del coseno del ángulo formado por el eje  $\alpha$ -ésimo y el vector  $pf_i - G_f$ . De la misma manera para el caso de los perfiles columna, la contribución relativa del eje  $\alpha$ -ésimo al perfil  $pc_j$  es:

$$\text{corr}^2(\theta_{j\alpha}) = \frac{\|\text{Pr}_{v_\alpha}(pc_j)\|_{D_p^{-1}}^2}{\|(pc_j - G_c)\|_{D_p^{-1}}^2} = \frac{[\text{coord}_{v_\alpha}(pc_j)]^2}{\sum_{i=1}^n \frac{T}{f_i} \left( \frac{k_{ij}}{c_j} - \frac{f_i}{T} \right)^2}.$$

Cuanto más grande sea  $\text{corr}^2(\theta_{i\alpha})$ , más específico es el perfil  $pf_i$  del eje  $\alpha$ -ésimo. La misma relación vale para los perfiles columna. Por otra parte, cuanto mayor sea  $\text{corr}^2(\theta_{i\alpha}) + \text{corr}^2(\theta_{i\beta})$ , de mejor calidad es la representación del perfil  $pf_i$  en el plano determinado por  $u_\alpha$  y  $u_\beta$ . Lo mismo vale para los perfiles columna.

### 3.3. Selección de ejes

Los criterios usuales para seleccionar el número de ejes son de carácter empírico. Sean  $a = \min\{n, p\}$ ,  $IE(r) = 100 \frac{\sum_{s=2}^r \lambda_s}{\sum_{s=2}^a \lambda_s}$  la inercia explicada por los primeros  $r - 1$  ejes, y sea  $ie(r) = 100 \frac{\lambda_r}{\sum_{s=2}^a \lambda_s}$  la inercia explicada por el  $r$ -eje. Una primera forma (sencilla) es fijar a priori un porcentaje de inercia explicada por los ejes, digamos por ejemplo 75 %, y escoger los primeros  $r - 1$  ejes tales que  $IE(r) \geq 75$ . Al ser este un criterio global, se aconseja controlarlo con un criterio local que involucre el porcentaje de inercia explicada por un eje:  $ie(r) \geq \frac{100}{a-1}$ . El criterio local consiste en retener los primeros  $r$  ejes tales que  $ie(r) \geq \frac{1}{a-1} \sum_{s=2}^a ie(s) = \frac{100}{a-1}$ , y  $ie(r+1) < \frac{100}{a-1}$ .

Otro criterio empírico, también usado en ACP, es la “regla del codo” que consiste en construir un diagrama lineal de los valores propios y determinar el punto donde la curva tiene una forma similar a un codo. Esto es, escoger los primeros  $r$  ejes tales que a partir del valor propio  $\lambda_r$ , el diagrama es aproximadamente una función constante.

### 3.4. Selección de puntos

La media aritmética de las contribuciones absolutas de los perfiles fila ( resp. perfiles columna) es  $\frac{1}{n}$  ( resp.  $\frac{1}{p}$ ), entonces los perfiles tales que  $\text{ctr}_\alpha(i) \geq \frac{1}{n}$  y  $\text{ctr}_\alpha(j) \geq \frac{1}{p}$  se llaman **perfiles explicativos** del eje  $\alpha$ -ésimo. En la etapa de depuración e interpretación de resultados se tomarán en cuenta prioritariamente los perfiles explicativos.

**Selección de puntos explicativos:** para el eje  $\alpha$ -ésimo supongamos que las contribuciones  $\text{ctr}_\alpha(i)$  están ordenadas en forma decreciente. Se escogen los  $h$  primeros puntos explicativos tales que  $\sum_{i=1}^h \text{ctr}_\alpha(i) \geq d$ , donde  $d$  es un número entre cero y uno, escogido a priori. El criterio para los perfiles columna es igual:  $\sum_{j=1}^g \text{ctr}_\alpha(j) \geq d$ .

Por otra parte, los perfiles fuertemente asociados con un eje se llaman puntos **explicados** por este eje. Normalmente se toma 0.5 como valor límite. Esto significa que un perfil  $pf_i$  es explicado por el eje  $\alpha$ -ésimo, si  $\text{corr}^2(\theta_{i\alpha}) \geq 0,5$ . En modo análogo, un perfil  $pc_j$  es explicado por el eje  $\alpha$ -ésimo, si  $\text{corr}^2(\theta_{j\alpha}) \geq 0,5$ .

Eventualmente un eje que explica muy poca inercia (no pasa el criterio de selección de ejes) puede ser considerado dentro del análisis si existe algún perfil **explicado** por este eje, de modo tal que se pueda afirmar que se trata de una dirección característica de ese perfil.

## 4. El algoritmo del análisis factorial de correspondencias

**Paso 1. Entrada de datos:** Los datos de entrada se presentan bajo el formato de una matriz  $K = (k_{ij})_{n \times p}$  con las siguientes propiedades:

- (a)  $k_{ij} \geq 0$ ;  $i = 1, \dots, n$ ;  $j = 1, \dots, p$ .
- (b) Se puede sumar por filas y columnas de  $K$ .
- (c)  $\sum_{j=1}^p k_{ij} > 0$  para todo  $i$ ,  $\sum_{i=1}^n k_{ij} > 0$  para todo  $j$ .

## Paso 2. Calcular la matriz a diagonalizar

La matriz a diagonalizar es  $Z = (z_{jl})_{a \times a}$  tal que

$$z_{jl} = \begin{cases} \frac{1}{\sqrt{c_j c_l}} \sum_{i=1}^n \frac{k_{ij} k_{il}}{f_i} & \text{si } p \leq n \\ \frac{1}{\sqrt{f_j f_l}} \sum_{i=1}^p \frac{k_{ji} k_{li}}{c_i} & \text{si } p > n \end{cases}$$

donde:  $a = \min \{p, n\}$ ;  $f_i = \sum_{h=1}^p k_{ih}$  es el total de la fila  $i$  de  $K$  y  $c_j = \sum_{t=1}^n k_{tj}$  es el total de la columna  $j$  de  $K$ .

## Paso 3. Cálculo de coordenadas

### Paso 3.0 Calcular los valores y vectores propios de $Z$

Denotamos con  $w_0, w_1, \dots, w_a$  los vectores propios de  $Z$ ,  $I_a$ -ortonormados, asociados a los valores propios  $1 > \lambda_2 \geq \dots \geq \lambda_a > 0$ . Entonces se tienen dos casos:

#### Paso 3.1 Primer caso: $p \leq n$

Para  $i = 1, \dots, n$  y  $\alpha = 1, \dots, a$  se calculan las coordenadas de los perfiles fila mediante la fórmula:  $\text{coord}_{u_\alpha}(pf_i) = pf_i D_q^{-\frac{1}{2}} w_\alpha$ . Usando las formas explícitas de  $pf_i = \left(\frac{k_{i1}}{f_i}, \dots, \frac{k_{ip}}{f_i}\right)$ ,  $D_q^{-\frac{1}{2}} = \text{diag}\left(\sqrt{\frac{T}{c_j}}\right)_{p \times p}$  y  $w_\alpha^t = (w_{\alpha 1}, \dots, w_{\alpha p})$ ; obtenemos una expresión para las coordenadas de los perfiles fila dependiendo de los  $w_{\alpha j}$  ( $i = 1, \dots, n$ ,  $\alpha = 2, \dots, a$ ):

$$\text{coord}_{u_\alpha}(pf_i) = \frac{\sqrt{T}}{f_i} \sum_{j=1}^p \frac{k_{ij} w_{\alpha j}}{\sqrt{c_j}}.$$

Para calcular las coordenadas de los perfiles columna se usan las fórmulas baricéntricas ( $j = 1, \dots, p$ ;  $\alpha = 2, \dots, a$ ) calculándose:

$$\text{coord}_{v_\alpha}(pc_j) = \frac{1}{c_j \sqrt{\lambda_\alpha}} \sum_{i=1}^n k_{ij} \text{coord}_{u_\alpha}(pf_i)$$

#### Paso 3.2 Segundo caso: $p > n$

Para  $j = 1, \dots, p$  y  $\alpha = 2, \dots, a$  se calculan las coordenadas de los perfiles columna mediante la fórmula:  $\text{coord}_{v_\alpha}(pc_j) = pc_j D_p^{-\frac{1}{2}} w_\alpha$ . Análogamente al caso anterior, usando las expresiones de  $pc_j^t = \left(\frac{k_{1j}}{c_j}, \dots, \frac{k_{nj}}{c_j}\right)$ ,  $D_p^{-\frac{1}{2}} = \text{diag}\left(\sqrt{\frac{T}{f_i}}\right)_{n \times n}$  y de  $w_\alpha^t = (w_{\alpha 1}, \dots, w_{\alpha n})$  se llega a:

$$\text{coord}_{v_\alpha}(pc_j) = \frac{\sqrt{T}}{c_j} \sum_{i=1}^n \frac{k_{ij} w_{\alpha i}}{\sqrt{f_i}}$$

Las coordenadas de los perfiles fila se calculan por medio de las fórmulas baricéntricas:

$$\text{coord}_{u_\alpha}(pf_i) = \frac{1}{f_i \sqrt{\lambda_\alpha}} \sum_{j=1}^p k_{ij} \text{coord}_{v_\alpha}(pc_j)$$

#### Paso 4. Ayudas a la interpretación

**Paso 4.1** Para  $i = 1, \dots, n$  y  $\alpha = 2, \dots, a$

**Paso 4.1.1** Calcular:

$$\text{ctr}_\alpha(i) = \frac{f_i [\text{coord}_{u_\alpha}(pf_i)]^2}{T \lambda_\alpha} \text{ y } \text{corr}^2(\theta_{i\alpha}) = \frac{[\text{coord}_{u_\alpha}(pf_i)]^2}{\sum_{j=1}^p \frac{T}{c_j} \left( \frac{k_{ij}}{f_i} - \frac{c_j}{T} \right)^2}$$

**Paso 4.1.2** Para  $\alpha = 2, \dots, a$ :

- (a) Crear una lista ordenada de las etiquetas de los perfiles fila según el orden decreciente de  $\text{corr}^2(\theta_{i\alpha})$ .
- (b) Crear una lista ordenada de las etiquetas de los perfiles fila según el orden decreciente de  $\text{ctr}_\alpha(i)$ .

**Paso 4.2** Para  $j = 1, \dots, p$  y  $\alpha = 2, \dots, r$

**Paso 4.2.1** Calcular:

$$\text{ctr}_\alpha(j) = \frac{c_j [\text{coord}_{v_\alpha}(pc_j)]^2}{T \lambda_\alpha} \text{ y } \text{corr}^2(\theta_{j\alpha}) = \frac{[\text{coord}_{v_\alpha}(pc_j)]^2}{\sum_{i=1}^n \frac{T}{f_i} \left( \frac{k_{ij}}{c_j} - \frac{f_i}{T} \right)^2}$$

**Paso 4.2.2** Para  $\alpha = 1, \dots, r$ :

- (a) Crear una lista lista ordenada de las etiquetas de los perfiles columna según el orden decreciente de  $\text{corr}^2(\theta_{j\alpha})$ .
- (b) Crear una lista lista ordenada de las etiquetas de los perfiles columna según el orden decreciente de  $\text{ctr}_\alpha(j)$ .

**Paso 5. Representación en el espacio bidimensional:** para hacer las representaciones de los perfiles fila, de los perfiles columna y de las representaciones simultáneas, el usuario selecciona los planos principales que desea. Dado  $\gamma \in ]0, 1[$  ( $\gamma$  suministrado por el usuario):

**Paso 5.1** Escoger desde la listas creadas en 4.1.2 los perfiles fila tales que  $\text{corr}^2(\theta_{i\alpha}) \geq \gamma$ , los cuales serán representados en los planos principales. Las coordenadas se calculan según las fórmulas 3.1 ó 3.2.

**Paso 5.2** Escoger desde la lista creada en 4.2.2 los perfiles columna tales que  $\text{coor}^2(\theta_{j\alpha}) \geq \gamma$ , los cuales serán representados en los planos principales. Las coordenadas se calculan según las fórmulas 3.1 ó 3.2.

**Paso 5.3** Seleccionar perfiles y fila y columna simultáneamente para su representación, siguiendo el mismo procedimiento que en 5.1 y 5.2.

## 5. Algunos detalles sobre la implementación

El algoritmo para el Análisis Factorial de Correspondencias presentado en la sección anterior fue implementado en lenguaje C++ [7] como un módulo más del sistema PIMAD 2.1 [6]. Esto permitió aprovechar el núcleo del sistema PIMAD para efectuar los Análisis en Componentes Principales que son necesarios para el Análisis Factorial de Correspondencias y para generar el gráfico del Plano Principal.

Igual que los demás módulos de PIMAD, este módulo debe ser ejecutado bajo Windows 3.1 o Windows95.

El módulo de Análisis Factorial de Correspondencias está implementado de modo tal que puede ser ejecutado con tablas de datos de cualquier tamaño, limitado solamente por la cantidad de memoria del computador en donde se esté ejecutando.

Figura 1: Interfaz del módulo AFC en PIMAD

En la Figura 1 se muestra la interfaz del sistema. Nótese que los cálculos se pueden realizar en forma directa a través de la barra de herramientas (tool-bar), o bien a través del menú principal.

Figura 2: Menú AFC-Paso-a-Paso en PIMAD

El algoritmo de la sección anterior también se puede ejecutar paso por paso mediante el submenú **AFC-Paso-a-Paso**, que se muestra en la Figura 2. Todos los archivos creados por el programa son tipo ASCII.

Tal como se muestra en la Figura 2 mediante la opción **Calcular la matriz Z** el programa ejecuta el paso 2 del algoritmo presentado en la sección anterior, dejando la matriz  $Z$  en el archivo **MATRIZ\_Z.TXT**. Con la opción **Calcular los Vectores y Valores Propios** se puede ejecutar el paso 3.0 del algoritmo, los valores propios quedarán almacenados en el archivo **VALORP.TXT** y los vectores propios en el archivo **VECTORP.TXT**. Con la opción **Calcular Coordenadas** se pueden ejecutar los pasos 3.1 y 3.2 del algoritmo, las coordenadas de los perfiles fila y columna quedan en el archivo **COORDENA.TXT**. Con la opción **Graficar el Plano Principal...** se despliega por pantalla el plano principal, tal como se ilustra más abajo (Figura 3.).

Una vez graficado el plano principal, mediante las opciones **Calcular Contribuciones Absolutas** y **Calcular Contribuciones Relativas** se puede ejecutar el paso 4 del algoritmo de la sección anterior. Las contribuciones absolutas se almacenan en el archivo **C\_ABSOLU.TXT** y las contribuciones relativas en el archivo **C\_RELATI.TXT**.

## 6. Ejemplo de ilustración

En esta sección se ilustra el uso del submenú **AFC-Paso-a-Paso**. La tabla de datos utilizada es de tamaño  $50 \times 5$  y proviene del “cruzamiento” de las variables *unidad académica de la UCR* con *actividad académica*. La casilla  $(i, j)$  contiene el número de tiempos completos invertidos por la unidad académica  $i$ , en la actividad académica  $j$ , en el II ciclo académico de 1993. Es claro que dicha tabla de datos no negativos, es sumable por filas y columnas por lo que se pueden estudiar las relaciones entre las modalidades por medio del AFC. Las unidades académicas son codificadas con los números del 1 al 50, y las actividades con abreviaciones cuyo significado se indica en el cuadro siguiente:

<i>Abreviación</i>	<i>Nombre de la Actividad</i>
Docencia	Actividad Docente
Invest	Trabajo de Investigación
ComInst	Trabajo en Comisiones Institucionales
DocAdm	Trabajo Docente-Administrativo
AcSoc	Acción Social

Las “salidas” correspondientes a cada una de las opciones del submenú “AFC paso-a-paso”, son las siguientes:

1. Opción: **Calcular la matriz Z**. El programa calcula la matriz Z de tamaño  $5 \times 5$  y crea el archivo `MATRIZ.Z.TXT` que tiene una forma similar a:

$$Z = \begin{matrix} & \begin{matrix} 0.719 & 0.309 & 0.157 & 0.190 & 0.144 \\ 0.309 & 0.162 & 0.081 & 0.114 & 0.077 \\ 0.157 & 0.081 & 0.072 & 0.081 & 0.046 \\ 0.190 & 0.114 & 0.081 & 0.193 & 0.074 \\ 0.144 & 0.077 & 0.046 & 0.074 & 0.056 \end{matrix} \end{matrix}$$

2. Opción: **Calcular los Vectores y Valores Propios**. Los vectores propios de Z son las columnas del archivo `VECTORP.TXT` el cual tiene una apariencia similar a:

$$\begin{matrix} & \begin{matrix} u_1 & u_2 & u_3 & u_4 & u_5 \\ -0.831 & 0.425 & 0.243 & 0.230 & 0.131 \\ -0.383 & -0.062 & -0.261 & -0.657 & -0.592 \\ -0.209 & -0.224 & -0.779 & 0.538 & -0.095 \\ -0.288 & -0.854 & 0.413 & 0.123 & -0.043 \\ -0.191 & -0.191 & -0.307 & -0.459 & 0.789 \end{matrix} \end{matrix}$$

Por otra parte los valores propios almacenados en `VALORP.TXT`, son:

$$\begin{matrix} \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 & \lambda_5 \\ 1 & 0.144 & 0.026 & 0.020 & 0.013 \end{matrix}$$

3. Opción: **Calcular Coordenadas**. El programa calcula las coordenadas de los perfiles fila y columna y crea un archivo de nombre `COORDENA.TXT`. En este archivo las 50 primeras filas son las coordenadas de las 50 Unidades Académicas (perfiles fila), y las últimas 5 son las coordenadas de las 5 actividades académicas (perfiles columna). Las siguientes tablas son segmentos de este archivo.

Coordenadas de las primeras 10 UA				Coordenadas de las actividades académicas			
Eje 1	Eje 2	Eje 3	Eje 4	Eje 1	Eje 2	Eje 3	Eje 4
0.351	0.101	0.040	-0.057	0.194	0.047	0.038	0.018
0.302	-0.011	-0.344	-0.374	-0.061	-0.109	-0.239	-0.173
0.403	0.099	0.118	0.000	-0.408	-0.599	0,358	-0,051
0.326	0.119	0.006	-0.053	-1.124	0,230	0.060	-0.017
0.027	-0.024	-0.009	0.048	-0,381	-0,259	-0,335	0,464
0.157	-0.096	-0.253	-0.093				
0.261	-0.052	0.224	0.153				
-1.126	0.309	0.194	-0.138				
-0.521	0.031	0.122	-0.052				
-0.585	-0.059	0.089	-0.026				

- Opción: **Graficar el Plano Principal**. El programa genera un plano principal en el cual el usuario puede elegir los perfiles y los ejes. El siguiente gráfico corresponde al plano de los ejes 1 y 2. Este gráfico puede ser impreso directamente con el programa, o bien el programa puede generar código L<sup>A</sup>T<sub>E</sub>X de este gráfico.
- Opción: **Calcular Contribuciones Absolutas**. Las contribuciones absolutas de las primeras 10 Unidades Académicas y las contribuciones absolutas de las actividades académicas (primeros 10 registros y los últimos 5 respectivamente, del archivo C\_ABSOLU.TXT), son:

Contribuciones Absolutas de las primeras 10 UA				Contribuciones Absolutas de las actividades académicas					
	Eje 1	Eje 2	Eje 3	Eje 4		Eje 1	Eje 2	Eje 3	Eje 4
1	4.460	2.0745	0.436	1.3623	Doc	18.068	5.907	5.279	1.708
2	0.357	0.003	3.463	6.273	DocAdm	0.383	6.710	43.156	35.018
3	2.862	0.969	1.830	0.000	ComIns	5.030	60.752	28.947	0.907
4	1.167	0.865	0.003	0.358	Inv	72.875	17.093	1.528	0.188
5	0.013	0.056	0.010	0.473	AcSoc	3.646	9.449	21.095	62.179
6	0.193	0.400	3.734	0.777					
7	1.618	0.356	8.854	6.363					
8	17.497	7.401	3.892	3.032					
9	4.543	0.092	1.867	0.523					
10	2.185	0.125	0.378	0.051					

- Opción: **Calcular Contribuciones Relativas**. Las contribuciones relativas de las primeras 10 Unidades Académicas y las contribuciones realtivas de las actividades académicas (primeros 10 registros y los últimos 5 respectivamente, del archivo C\_RELATI.TXT),

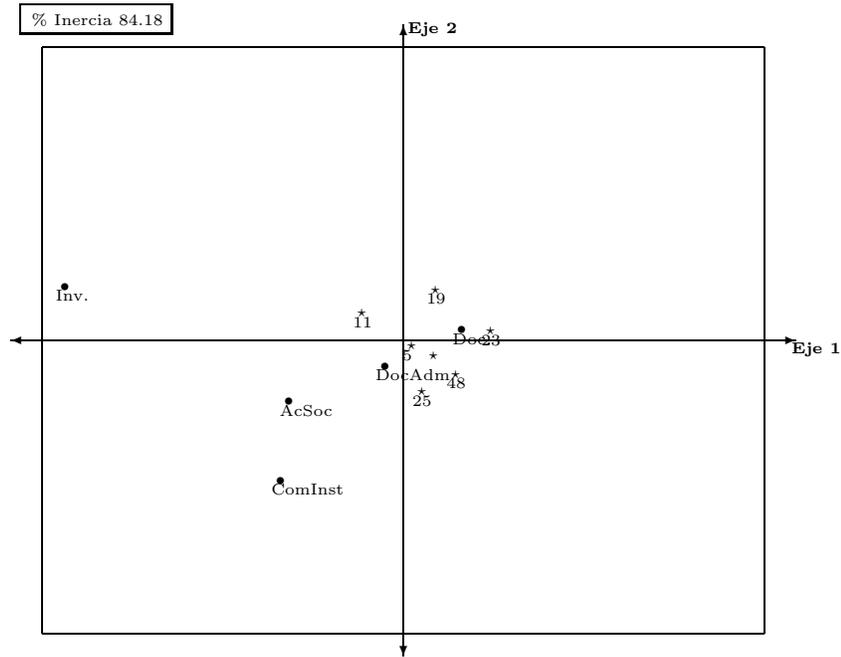


Figura 3: Plano Principal

son

Contribuciones relativas de las primeras 10 UA					Contribuciones relativas de las actividades académicas				
					Eje 1	Eje 2	Eje 3	Eje 4	
1	0.891	0.074	0.012	0.024	Doc	0.904	0.053	0.036	0.007
2	0.261	0.000	0.340	0.399	DocAdm	0.037	0.116	0.555	0.292
3	0.872	0.053	0.075	0.000	ComInst	0.253	0.547	0.196	0.004
4	0.862	0.114	0.000	0.023	Inv	0.957	0.040	0.003	0.000
5	0.200	0.152	0.020	0.627	AcSoc	0.268	0.124	0.209	0.399
6	0.232	0.086	0.601	0.081					
7	0.472	0.019	0.347	0.162					
8	0.892	0.067	0.027	0.014					
9	0.936	0.003	0.052	0.009					
10	0.966	0.010	0.022	0.002					

## Referencias

- [1] Benzécri, J.P. (1973) *L'Analyse des Données. Tomo 2: L'Analyse des Correspondances.*, (2<sup>de</sup>. éd. 1976). Dunod, París.
- [2] Cailliez, F.; Pagès, J. P. (1976) *Introduction à l'Analyse des Données.* SMASH, París.
- [3] Castillo, W. (1997) *Análisis Factorial de Correspondencias: Método, teoría y algoritmos.* Programa de Investigación en Modelos y Análisis de Datos (PIMAD), Escuela de Matemática, Universidad de Costa Rica, San José.
- [4] Diday, E.; Lemaire, J.; Pouget, J.; Testu, F. (1982) *Eléments d'Analyse de Données.* Dunod, París.
- [5] Jambu, M. (1989) *Exploration Informatique et Statistique des Données.* Dunod, París.
- [6] Lebart, L.; Morineau, A.; Piron, M. (1995) *Statistique Exploratoire Multidimensionnelle.* Dunod, París.
- [7] Rodríguez, O. (1996) *PIMAD 2.1 Manual al Usuario.* Programa de Investigación en Modelos y Análisis de Datos, Universidad de Costa Rica, San José.
- [8] Rodríguez, O. (1997) *Introducción a la Programación Orientada a Objetos en C++ para Ambiente Windows.* Editorial Tecnológica de Costa Rica, Cartago.