15
2 ej-

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

Facultad de Ciencias

UN PROGRAMA PARA REALIZAR ANALISIS FACTORIAL DE CORRESPONDENCIAS

T E S I S

Que para obtener el grado de Licenciatura en
M A T E M A T I C A S

p r e s e n t a

MA. ALMA GARCIA GARCIA

México, D. F. Noviembre de 1992

ESQUEMA

INTRODUCCION

El análisis de Correspondencias es un tema relativamente nuevo y por consiguiente aún en desarrollo. Uno de los pioneros en esta área es Jean-Paul Benzécri en Francia. Dicha técnica fue usada primeramente (y de hecho ahí comenzó el desarrollo de la teoría) en Psicología al querer clasificar una gran cantidad de elementos con respecto a un numeroso conjunto de variables*.

Interpretar la masa de datos que se tenía no era fácil y tanto más éstos crecieran, mayor sería esa dificultad, sobre todo gráficamente, y ni mencionar los cálculos que se debían hacer para ellas.

En cuanto a las formas gráficas de representación se usaban los dendogramas, en ocasiones bastante complejos de descifrar para alguien con pocos conocimientos sobre el tema, es así como fue necesario un método de descripción sintética de matrices de grandes dimensiones con una gran cantidad de datos numéricos. Lo anterior implica hablar de estadística descriptiva multidimensional como: diagramas triangulares, estadística multivariada inferencial, análisis de componentes principales, análisis canónico, etc.

El problema que existía era el tamaño de las tablas, porque hacer las operaciones aritméticas y algebráicas manualmente, tomaba demasiado tiempo. Afortunadamente sobrevino el creciente desarrollo de las computadoras que aunque por la década de los 60's no eran tan veloces en sus cálculos como ahora, ayudaron en gran medida a que las operaciones necesarias de la tabla de datos fueran más rápidas, y sobre todo, cuando era tan vasta la cantidad de información. Y no sólo la computadora operaba con los datos, sino que también las técnicas de graficación han

*. Székely, Béla, *Los test*, p. 153.

ayudado a la interpretación de los resultados por el mismo medio, además con la tecnología actual, la multimedia permite la presentación de los datos y las gráficas de tal manera que su manejo sea muy sencillo.

El objetivo de este trabajo es desarrollar un programa que realice análisis factorial de correspondencias, mostrando los resultados en forma de tablas y gráficas. A lo largo del texto se describirá el método utilizado y las técnicas empleadas, así como algunos ejemplos de su aplicación en la siguiente secuencia: en el Capítulo I se hace una pequeña presentación de los datos que serán tratados en este trabajo y en el programa; también se muestra algunas formas de transcripción de ellos. Finalmente se introduce y explica el término de elementos suplementarios.

El Capítulo II describe los elementos geométricos y algebráicos que son utilizados en el texto. A continuación se trata brevemente el análisis canónico y el análisis en componentes principales, ya que ambos, proporcionan técnicas al análisis factorial de correspondencias.

En el Capítulo III se explica el programa que realiza el análisis factorial de correspondencias por medio del seguimiento de un ejemplo.

El Capítulo IV está formado con el Manual de Uso.

Con el Capítulo V se muestra la utilización del programa con algunos ejemplos.

Finalmente se exponen las conclusiones surgidas de esta obra.

I. PRESENTACION

I.1 Introducción a las variables cualitativas y cuantitativas

La estadística clásica es representada en los textos usuales como el estudio de una sola variable que es medida sobre un conjunto de individuos, el valor que toma el individuo con respecto a una variable es de acuerdo a dos criterios: cualitativo y cuantitativo.

Cuando sea cuantitativamente, el individuo tomará valores de una escala numérica que además puede ser: a) de razón: cuando ordenados los datos se pueda indicar la distancia exacta entre ellos y además se tenga la posibilidad de situar al cero absoluto. b) de intervalo: cuando ordenados los objetos sea factible determinar la distancia entre ellos. Entre los casos anteriores se considera al salario, edad, tallas, etc.

Pero si se habla de variables cualitativas, la información cambia: los datos no son numéricos, sino que están divididos en categorías o niveles que se procura estén completos para que todos los individuos pertenezcan a alguna de ellas, y además las categorías deben ser mutuamente excluyentes para con esto evitar que exista la pertenencia a dos o más de ellas (por ejemplo: sexo {femenino, masculino}, religión {budista, cristiana, judía, etc.}). En la práctica se observa que los individuos son medidos por más de una variable y las variables suelen tener varias categorías (en el caso de variables cualitativas que son las más comunes), por lo que normalmente se hablará de análisis multivariado.

Los datos que son las observaciones de las variables en los individuos, son representados usualmente mediante tablas, a continuación se describen las diferentes tablas que se usarán a lo largo del texto así como en el programa.

I.1.1 Matriz de individuos por variables

Al hablar de una matriz se refiere por supuesto a una tabla de doble entrada en la que las líneas representan a los individuos y las columnas a las variables:

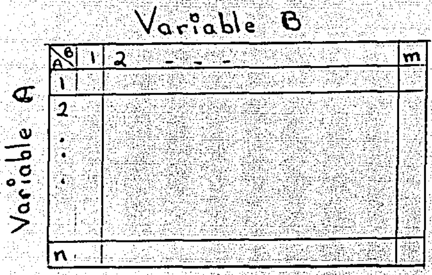
$$
\begin{array}{cc}
 & \text{variables} \\
\text{individuos} & \begin{pmatrix}
x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1p} \\
x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2p} \\
\cdot & \cdot & & \cdot & & \cdot \\
x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{ip} \\
\cdot & \cdot & & \cdot & & \cdot \\
x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{np}
\end{pmatrix} = X_{np}
\end{array}
$$

la tabla anterior es apropiada cuando los datos que se tienen son numéricos o cuantitativos. Se observa que *p* características cuantitativas se miden sobre *n* individuos, es decir: el individuo *i* toma los valores $x_{i1}, x_{i2}, \dots, x_{ip}$ respectivamente de las *p* variables, o dicho de otra forma: los *n* individuos toman respectivamente a $x_{1j}, x_{2j}, \dots, x_{nj}$ como valores de la variable *j*.

I.1.2 Tablas de contingencia

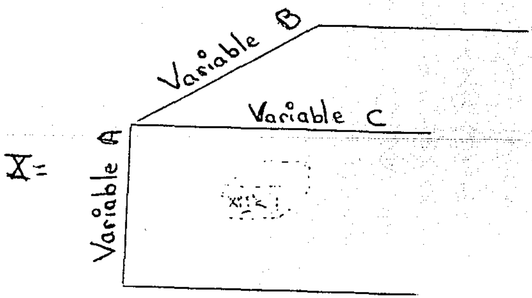
Las tablas de contingencia son utilizadas cuando las variables que se analizan son cualitativas, esta forma de presentación de los datos muestra las interrelaciones entre las categorías de las variables. Por ejemplo sexo contra posición política o religión contra preferencias gastronómicas.

En la siguiente figura puede apreciarse un esquema de una tabla de contingencia con dos variables:

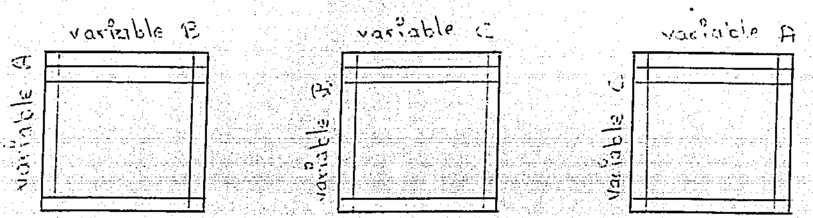


se observa que los renglones de la tabla corresponden a las categorías de la variable A y las columnas a las categorías de la variable B.

El ejemplo anterior fue presentado con dos variables por ser un caso trivial, sin embargo, el análisis puede complicarse con 3, 4, 5 ó más variables: con tres variables puede formarse un cubo como el de la siguiente figura:



la cual indica que en la celda (i,j,k) se encuentra el número de individuos que pertenecen a la categoría *i* de la variable A, a la categoría *j* de la variable B y a la categoría *k* de la variable C. Como puede apreciarse, mientras más grande sea el número de variables que intervienen en el análisis, más difícil será su interpretación geométrica, obviamente más de tres es imposible graficarla aunque se conozca que se formaría un hipercubo, pero, lo que sí puede realizarse es la descomposición de los hipercubos en tablas, con la composición de variables que se deseen, por ejemplo:



con el desglosamiento del hipercubo en tablas resulta más fácil observar algunas relaciones, pero no en conjunto, es decir, que si se quisiera visualizar las relaciones entre tres variables, el cubo sería muy difícilmente observable. Y ni qué decir de los hipercubos.

I.1.3 Tablas de variables indicatrices o matrices disyuntivas completas

El tipo de tablas que a continuación se presenta es de descripción lógica o de forma booleana donde los valores son exclusivamente de ceros y unos. Esta representación de información es muy útil para las variables cualitativas pues con ello se indica si el individuo observado pertenece o no a una categoría de alguna variable. A continuación se muestra el esquema de una tabla con variables indicatrices:

i \ j	1-2	3-4	. . .	50-51	. . .	99-100
1	0	0		0		0
2	1	0		0		0
3	0	0		1		0
n	0	1		0		0

donde:

líneas = individuo identificado con el número i donde $i = 1, \dots, n$

columnas = edades 1-2, 3-4, ... 99-100 años (50 categorías)

$$a_{ij} = \begin{cases} 1 & \text{si el } i\text{-ésimo individuo pertenece a la categoría } j \\ 0 & \text{en otro caso} \end{cases}$$

Casi cualquier tabla numérica puede ser expresada como una tabla compactada de una tabla de variables indicatrices:

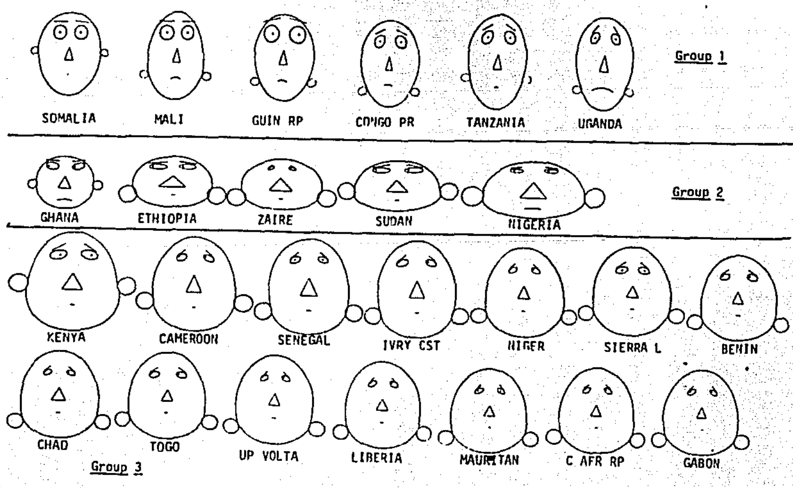
	fuma	no-fuma
director	0	1
subdirector	1	0
secretaria 1	1	0
secretaria 2	1	0
secretaria 3	1	0
empleado 1	0	1
empleado 2	0	1
empleado 3	1	0

	fuma	no-fuma
dirección	1	1
secretarias	3	0
empleados	1	2

I.2 Transcripción de datos

Después de haber expuesto las tablas de datos que se utilizarán a lo largo del trabajo y en el programa, se muestra a continuación la transcripción de éstas, pues es mucho más fácil observar las relaciones en una gráfica que en una tabla, de aquí la importancia del análisis factorial de correspondencias.

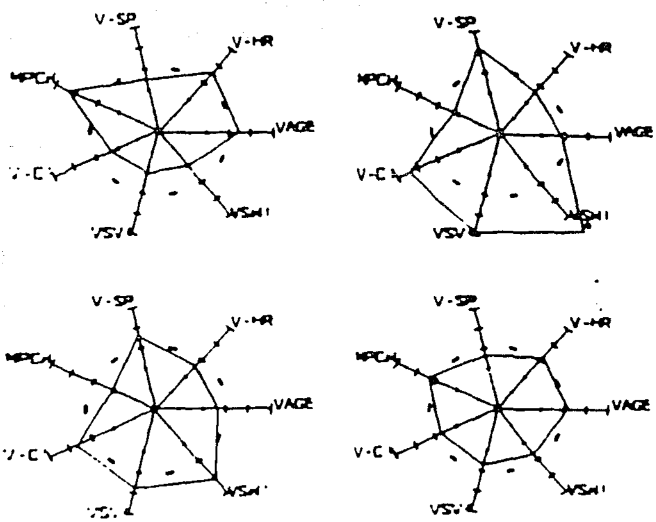
I.2.1 Chernoff, propuso un esquema de cara con los siguientes parámetros: boca, nariz, ojos, cejas, forma de cara*. Si los valores de los parámetros no fueran muchos sería muy fácil el observar las diferentes fisonomías, pero al crecer tanto los parámetros, como las categorías, las combinaciones que se forman dificultan la distinción a simple vista de las caras:



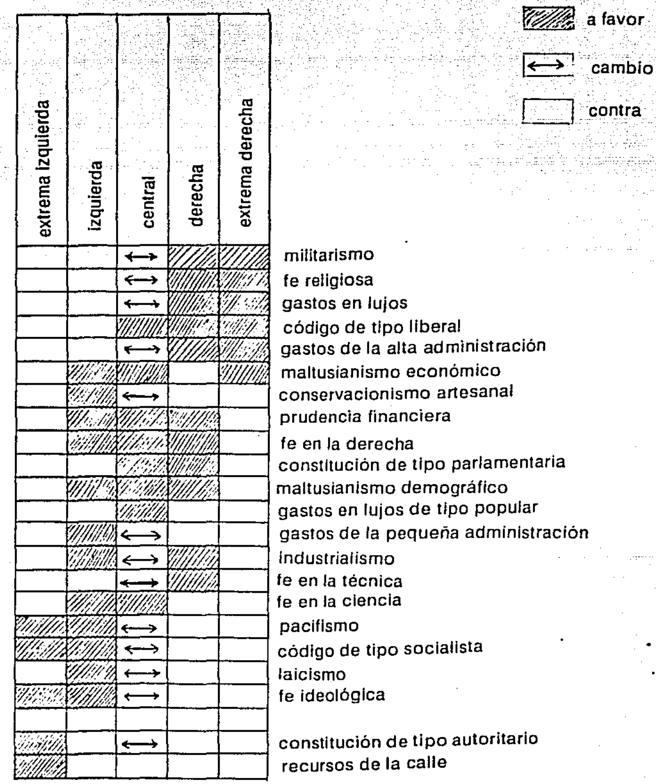
* Wang, Graphical Representation, p.39

1.2.2 El método de polígonos es más sistemático. Se trata de un vértice del cual salen varios ejes, cada uno de ellos representa a una variable. Cada eje tiene una marca que significa el valor que toma el individuo en esa variable, al unir todas las marcas se forma un polígono, y dependiendo del área de aplicación, es el uso que se da. Por ejemplo en Psicología, ciertos polígonos son representativos de varias desviaciones mentales. Para trazarlo se dan los siguientes pasos:

a) Sea m el número de parámetros.
b) Se divide el círculo en m sectores con el mismo ángulo, marcando m puntos en el círculo.
c) Se trazan radios a cada punto del círculo, marcando las mismas unidades a lo largo de todas y cada una de ellas.
d) El valor del parámetro es marcado en la línea que le corresponda.
e) Se unen las últimas marcas y se obtiene un polígono para cada valor de los parámetros.



I.2.3 El método gráfico o visual de Jacques Bertin es un método de tablas mucho más elaborado. Se trata de una tabla en la que cada caso es codificado con valores pero no numéricos, sino que son tonos grises que van del negro al blanco. Las líneas son los objetos y las columnas representan a los parámetros.



Ahora como es de suponerse, se tendrá una matriz de datos en la cual, a continuación se hará la aclaración de lo que representan las líneas y lo que representan las columnas.

Cuando se tomen tablas con variables indicatrices, se tomarán las líneas como individuos a estudiar, o sea, los objetos a los cuales se les hará el análisis, éstos pueden ser por ejemplo: países, clases de plantas, de animales, temas de libros, etc., a los cuales se les asocian ciertos parámetros o variables. Las columnas serán las variables, éstas calificarán cada individuo y pueden ser por ejemplo: características físicas, medidas o frecuencias, categorías profesionales, etc.

Sea entonces una tabla rectangular de datos con n líneas y p columnas, estos datos pueden ser representados como puntos en un espacio de n o de p dimensiones respectivamente. Las n líneas serán n puntos en el espacio R^p, y las p columnas serán p puntos en el espacio R^n.

Dados los puntos es posible evaluar las distancias elegidas de manera que describan las similitudes entre líneas, sin embargo, pueden visualizarse los puntos en hasta tres dimensiones, y aún aquí con un poco de dificultad; por lo que sería conveniente tener una imagen de los puntos en un plano, donde las propiedades de proximidad entre ellas no se deformen mucho, siendo así que la fórmula de distancia a utilizar (la métrica), el criterio para ajustar los puntos, y en sí, la geometría de los puntos, son variables y deben ser elegidos de tal manera que los resultados sean de lectura simple.

I.3 Elementos Suplementarios

Otro concepto que se está manejando (y que es otra posibilidad a usar dentro del programa), es el de los elementos suplementarios, que son: individuos y variables suplementarias, éstos nos indican que la información que proporcionan es puramente adicional a la que ya se tiene, ayudan para dar mayor precisión a los resultados.

En el caso de las variables suplementarias indican otros elementos que en cierto momento pueden relacionarse con las variables activas del análisis para aclarar la interpretación, por ejemplo véase la siguiente situación*:

En un grupo de cinco clases de trabajadores: 1) Gerencia Principal, 2) Subgerentes, 3) Empleados de edad madura, 4) Empleados jóvenes y 5) Secretarias; se tienen cuatro niveles de adicción a fumar: I) Abstemios, II) Baja, III) Mediana, IV) Alta.

Además existen dos variables suplementarias que son V) Ingieren alcohol y VI) No ingieren alcohol. La siguiente tabla muestra el ejemplo:

	Consumo de tabaco				Consumo de alcohol	
	(1) No	(2) Bajo	(3) Medio	(4) Alto	(1) No	(2) Sí
(1) Senior managers	4	2	3	2	0	11
(2) Junior managers	4	3	7	4	1	17
(3) Senior employees	25	10	12	4	5	46
(4) Junior employees	18	24	33	13	10	78
(5) Secretaries	10	6	7	2	7	18
Promedio	42%	29%	20%	9%		

* Greenacre, M. J., *Theory and applications of ...,* p 72.

Siguiendo el ejemplo anterior se tendrán ahora como individuos suplementarios a: 6) Personal que trabaja de día, 7) Personal que trabaja en la tarde y 8) Personal que trabaja de noche.

Consumo de tabaco

	Abstemios	Baja adicción	Mediana adicción	Alta adicción
1 Jefes de edad	4	2	3	2
2 Jefes jóvenes	4	3	7	4
3 Empleados de edad	25	10	12	4
4 Empleados jóvenes	18	24	33	13
5 Secretarias	10	6	7	2
6 Personal matutino	14	22	47	18
7 Personal vespertino	25	20	14	5
8 Personal nocturno	2	3	1	2

En cualquier momento pueden existir tanto individuos como variables suplementarias, debe entonces considerarse lo siguiente:

— los elementos suplementarios tienen peso nulo o masa cero.

— no entran en el proceso de obtención de los valores propios y por lo tanto tampoco de los vectores propios.

— los elementos suplementarios saldrán impresos en tablas separadas de los elementos activos.

— las gráficas pueden salir para cada tipo de elementos o para ambos: activos y suplementarios.

Como se considerará a las variables suplementarias con peso nulo, no hay problema al tratarlas como a las demás en lo que resta de los cálculos, sin embargo, se mostrará cómo se utilizan en el programa posteriormente.

II. ELEMENTOS GEOMETRICOS Y ALGEBRAICOS

II.1 Algunas definiciones

En esta sección se dará la notación y algunas definiciones que se usarán en los capítulos siguientes.

Como los datos son presentados en una matriz, es muy fácil manejarlos en forma de vectores (columna) así, a cada individuo se le asocia un vector en el espacio vectorial R^p.

$$\bar{x}_i = \begin{pmatrix} x_{1i} \\ x_{2i} \\ \cdot \\ x_{ji} \\ \cdot \\ x_{pn} \end{pmatrix} \in R^p$$

de igual manera cada variable tiene asociado un vector en el espacio vectorial R^n.

$$\bar{y}_j = \begin{pmatrix} y_{j1} \\ y_{j2} \\ \cdot \\ y_{jm} \\ \cdot \\ y_{jn} \end{pmatrix} \in R^n$$

Se llamará M a la nube de puntos que forman los n-individuos, esto es, M= $\{x_i\}_{i \in I}$ y

N a la nube formada por las variables: N = $\{x_j\}_{j \in J}$.

Al espacio de individuos se le denotará como E tal que $E \subset R^p$

Al espacio de variables se le nombrará F tal que $F \subset R^n$

E y F forman los espacios en los cuales, respectivamente M y N están inmersos.

A cada individuo se le dará un peso p_i tal que

$$p_i > 0 \ ; \qquad \sum_{i=1}^{n} p_i = 1$$

y como pueden existir elementos suplementarios, es necesario remarcar que éstos no pueden tener peso, ya que, dentro de las categorías, éstos entrarían doblemente en ellas y como se dijo, solamente pueden pertenecer a una (para garantizar la independencia).

II.2 Espacio de Individuos

Dentro de este espacio se darán algunas características, justamente para los individuos.

II.2.1 Centro de gravedad: Al tener todos los vectores de individuos en R^p (el espacio de variables), se puede obtener un centro de gravedad de los vectores formados, éste se define tomando el peso y las coordenadas como:

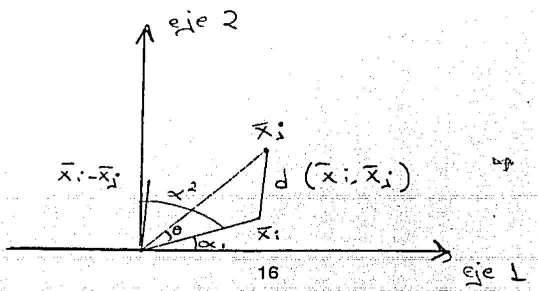
$$\bar{C} = \sum_{i=1}^{n} p_i x_i$$

esto indica que cada una de las coordenadas del centro de gravedad es un promedio ponderado de x_{ij} con $i = 1, \ldots, n$.

Como se está hablando de aspectos geométricos, es necesario dar una métrica que pondere cada variable de acuerdo a la importancia que tenga.

II.2.2 Métrica

La métrica será de la forma más general posible:



$$d^2(\bar{x}_i,\bar{x}_j) = ||x_i||^2 + ||x_j||^2 - 2||x_i||\,||x_j|| \cos\theta$$

donde

$$||x_i|| = (x^2_{i1} + x^2_{i2})^{1/2}$$

$$||x_j|| = (x^2_{j1} + x^2_{j2})^{1/2}$$

$$\cos\theta = \frac{(x_{i1}\cdot x_{j1} + x_{i2}\cdot x_{j2})}{((x^2_{i1} + x^2_{i2})(x^2_{j1} + x^2_{j2}))^{1/2}} = \frac{<\bar{x}_i, \bar{x}_j>}{<\bar{x}_i, \bar{x}_i> \; <\bar{x}_j, \bar{x}_j>}$$

asi que como producto de matrices se puede llegar a:

$$d^2(\bar{x}_i,\bar{x}_j) = ||x_i||^2 + ||x_j||^2 - 2||x_i||\,||x_j|| \cos\theta$$

$$= <\bar{x}_i, \bar{x}_i> + <\bar{x}_j, \bar{x}_j> - \frac{2(<\bar{x}_i, \bar{x}_i> \; <\bar{x}_j, \bar{x}_j>)^{1/2}}{(<\bar{x}_i, \bar{x}_i> \; <\bar{x}_j, \bar{x}_j>)^{1/2}} \cdot <\bar{x}_i, \bar{x}_j>$$

$$= \bar{x}^t_i \bar{x}_i + \bar{x}^t_j \bar{x}_j - 2\bar{x}^t_i \bar{x}_j$$

$$\therefore \quad d^2(\bar{x}_i, \bar{x}_j) = (\bar{x}_i - \bar{x}_j)^t (\bar{x}_i - x_j)$$

como se está usando un espacio ponderado es necesario dar una matriz que conserve esta característica, por lo que puede considerarse:

$$d^2(\bar{x}_i, \bar{x}_j) = \sum_j m_j (\bar{x}_i - \bar{x}_j)^2$$

$$d^2(\bar{x}_i, \bar{x}_j) = (\bar{x}_i - \bar{x}_j)^t \, \Phi \, (\bar{x}_i - \bar{x}_j)$$

donde se tiene que Φ determina la métrica del espacio de individuos. Por supuesto se debe verificar que efectivamente es una métrica, para esto Φ debe cumplir con que:

1) $d(x^1, x^2) = d(x^2, x^1)$

2) $d(x^1, x^2) = 0 \leftrightarrow x^1 = x^2$

3) $d(x^1, x^2) \geq 0$

por lo tanto Φ debe ser:

a) Simétrica: $m_{ij} = m_{ji}$

b) Definida: $\overline{x}^t \Phi \overline{x} = 0 \leftrightarrow x = 0$

c) Positiva: $\overline{x}^t \Phi \overline{x} > 0, \quad \forall x \neq 0$

se propone como matriz simétrica, positiva definida para métrica en el espacio de individuos a:

$$\Phi = \begin{pmatrix} 1/\varphi_1^2 & 0 & 0 & 0 & 0 \\ 0 & 1/\varphi_2^2 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & 1/\varphi_k^2 & \cdot \\ \cdot & \cdot & \cdot & 0 & \\ 0 & 0 & \cdot & 0 & 1/\varphi_p^2 \end{pmatrix}$$

donde

$$\varphi^2{}_k = \sum_{i=1}^{I} p_i \, (x_{ki} - x_{-k})^2$$

$$x_{-k} = \sum_{i=1}^{m} p_i \, x_{ki}$$

los elementos se pueden interpretar como los inversos de las varianzas y las medias de las variables $\overline{x}_k$ correspondientes. Se toma esta métrica, pues, tiene algunas ventajas muy interesantes:

$$d^2 (\overline{x}_i , \overline{x}_j) = \sum_{k=1}^{p} \frac{(x_{ki} - x_{kj})^2}{\varphi^2{}_k}$$

— Implicando con esto que si $\varphi^2{}_k$ es pequeña tendrá más importancia la diferencia de coordenadas y la nube de puntos será más regular, esto es, no tan dispersa (la dispersión se refiere a que algunos puntos se disparan en alguna dirección).

— No existe una unidad de referencia, pues cada individuo define la suya y, al tomar el conjunto de ellos no se pueden medir los otros con las unidades de alguno diferente.

— Considerando que toda matriz positiva definida puede expresarse como

$$X = T^t T$$

con T una matriz triangular superior*, entonces:

$$< \overline{x}_i , \overline{x}_j >_X = \overline{x}_i^t X \overline{x}_j = \overline{x}_i T^t T \overline{x}_j = < T\overline{x}_i , T\overline{x}_j >_I$$

y este caso, expresado por matrices:

$$M = \begin{pmatrix} 1/m_1^2 & & \\ & \cdot & \\ & & 1/m_p^2 \end{pmatrix} = \begin{pmatrix} 1/m_1 & & \\ & \cdot & \\ & & 1/m_p \end{pmatrix} \cdot \begin{pmatrix} 1/m_1 & & \\ & \cdot & \\ & & 1/m_p \end{pmatrix}$$

tomando $z_i = D_{1/m} x_i$, se trabajará con un nuevo espacio de individuos donde la distancia es la canónica, ya que, la métrica sería la matriz identidad (I).

$$d^2 (z_i, z_j) = < z_i, z_j >_I$$

La inercia de una nube con relación a un punto indica la dispersión de una respecto a un punto en R^p y se define como el promedio ponderado de los cuadrados de las distancias de los n-puntos al punto p:

$$I_p = \sum_{i=1}^{n} P_i || \overline{x}_i - \overline{p} ||^2_M$$

* $x_{ij} = x_{ji}$ porque para obtener x_{ij} se realizó la operación: $T^t_{ij} \cdot t_{ij}$ y para $x_{ji} = \overline{t}_{ij} \cdot T^t_{ij}$ para toda i, j = 1, ..., n por lo tanto X es simétrica y por definición (véase BURDEN, R. *Análisis Numérico*, p. 367) X es positiva definida.

En particular el punto p puede ser el centro de gravedad de la nube, en cuyo caso la fórmula es:

$$I = \sum_{i=1}^{n} P_i \| \bar{x}_i - \bar{g} \|^2_M$$

II.3 Espacio de variables

Así como en el espacio de individuos se definieron algunos elementos, en el espacio de variables también se definirán algunos conceptos.

Se puede dar un promedio ponderado de cada variable en el espacio de individuos:

$$(1) \qquad \bar{y}_j = \sum_{i=1}^{n} P_i \, y_{ji}$$

p_i es el peso asociado a cada individuo, la cantidad anterior forma la j-ésima coordenada del centro de gravedad de la nube en el espacio de individuos.

II.3.1 Centro de gravedad

En el caso de que $y_j = 0$, entonces la variable está centrada, permitiendo así trabajar con variables centradas haciendo una resta:

$$\sum_{i=1}^{n} P_i \, (x_{ij} - x_{-j}) = 0$$

obteniendo con esto que el centro de gravedad es el origen del sistema coordenado. Para la dispersión de una variable cualitativa se hace:

$$(2) \qquad S^2(\overline{y}_j) = \sum_{i=1}^{n} p_i (y_{ji} - y_{-j})^2$$

en caso de que los p_i sean probabilidades la fórmula (1) se toma como la media y la fórmula (2) como la varianza.

II.3.2 Métrica

Dentro del espacio de variables, también es necesario saber la proximidad de una con la otra, por lo que se necesita una métrica, ésta es definida por la matriz de pesos que es diagonal:

$$\Omega_p = \begin{pmatrix} \omega_1 & 0 & \cdot & 0 \\ 0 & \omega_2 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \omega_n \end{pmatrix}$$

el producto interno es:

$$\begin{aligned} < \overline{y}_j , \overline{y}_k >_{\Omega_p} &= \overline{y}_j^t \, \Omega_p \, \overline{y}_k \\ &= \sum_{i=1}^{n} p_i \, y_{ji} \, y_{kj} \end{aligned}$$

la norma

$$|| \overline{y}_j ||_{\Omega_p} = (< \overline{y}_j , \overline{y}_j >_{\Omega_p})^{1/2}$$

y la distancia entre dos puntos:

$$d^2(\overline{y}_j , \overline{y}_k) = (\overline{y}_j - \overline{y}_k)^t \, \Omega_p \, (\overline{y}_j - \overline{y}_k)$$

considerando que se pueden centrar las variables, la definición de producto interno y norma se pueden expresar como la covarianza y varianza respectivamente:

$$< \bar{y}_j , \bar{y}_k >_{\Omega_p} = \operatorname{cov}(\bar{y}_j , \bar{y}_k)$$

$$|| \bar{y}_j ||^2_{\Omega_p} = \operatorname{var}(\bar{y}_j)$$

expresando el coseno como producto interno:

$$\cos \theta = \frac{< \bar{y}_j , \bar{y}_k >_{\Omega_p}}{\sqrt{|| \bar{y}_j ||^2_{\Omega_p} \, || \bar{y}_k ||^2_{\Omega_p}}}$$

resulta que es el coeficiente de correlación entre y_j y y_k

Para comprender de manera amplia al análisis factorial de correspondencias, se revisará primero el método que engloba a éste y luego, a otro del cual se toman varias técnicas para la interpretación gráfica, éstos son: El Análisis Canónico y El Análisis de Componentes Principales respectivamente.

II.4 Análisis Canónico

H. Hotelling desde 1930 trabajaba con las técnicas de análisis de datos*, ocupándose por supuesto del análisis canónico, que como se ha mencionado es el método general de la regresión múltiple, del análisis de la varianza, del análisis de correspondencias, por ejemplo. Este método requiere considerablemente más trabajo de cálculos; es por esto que, la computadora proporciona una gran ayuda, así como en las demás técnicas. El objeto de estudio del análisis canónico es determinar hasta dónde se parecen dos paquetes de variables (con valores cuantitativos) independientes entre sí, bajo un conjunto de individuos, desde el punto de vista del análisis lineal. Sean las siguientes matrices de datos:

$$X_{n x p} = \begin{pmatrix} x_{11} & x_{12} & x_{1j} & x_{1p} \\ x_{21} & x_{22} & x_{2j} & x_{2p} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ x_{i1} & x_{i2} & x_{ij} & x_{ip} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ x_{n1} & x_{n2} & x_{nj} & x_{np} \end{pmatrix} \qquad Y_{n x q} = \begin{pmatrix} y_{11} & y_{12} & y_{1j} & y_{1q} \\ y_{21} & y_{22} & y_{2j} & y_{2q} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ y_{i1} & y_{i2} & y_{ij} & y_{iq} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ y_{n1} & y_{n2} & y_{nj} & y_{nq} \end{pmatrix}$$

en las cuales:

$$\vec{x_i} = \begin{pmatrix} x_{i1} \\ \cdot \\ \cdot \\ x_{ij} \\ \cdot \\ \cdot \\ x_{ip} \end{pmatrix} \in R^p$$

representa al i-ésimo individuo con los valores que toma cada una de las p-variables con respecto a él en el primer paquete.

* Greenacre, M. *Theory and Applications ....*, p. 108 ; BELA, S. *Los Tests,* p. 160

$$\overline{x}_j = \begin{pmatrix} x_{1j} \\ \cdot \\ \cdot \\ x_{ij} \\ \cdot \\ \cdot \\ x_{nj} \end{pmatrix} \in R^n$$

denota a la j-ésima variable con los valores que toman cada uno de los n-individuos con respecto a ella en el primer paquete.

Análogamente:

$$\overline{y_i} = \begin{pmatrix} y_{i1} \\ \cdot \\ \cdot \\ y_{ij} \\ \cdot \\ \cdot \\ y_{iq} \end{pmatrix} \in R^q$$

simboliza al i-ésimo individuo con los valores que toma cada una de las q-variables con respecto al individuo en el segundo paquete.

$$\overline{y}_j = \begin{pmatrix} y_{1j} \\ \cdot \\ \cdot \\ y_{ij} \\ \cdot \\ \cdot \\ y_{nj} \end{pmatrix} \in R^n$$

representa a la j-ésima variable con los valores que toman cada uno de los n-individuos con respecto a ella en el segundo paquete.

Nótese que las matrices conservan el mismo número de individuos, pues es con respecto a éstos que se realizará el análisis.

II.4.1 Métrica

Cada individuo tendrá un peso p_i tal que $p_i > 0$ y $\sum_{i=1}^{n} p_i = 1$. La métrica asociada al espacio de variables, será como se definió (positiva, definida, simétrica):

$$\Omega_{p \times p} = \begin{pmatrix} \omega_1 & 0 & 0 \\ 0 & \omega_2 & 0 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & 0 & \omega_n \end{pmatrix}$$

II.4.2 Centro de gravedad

Se va a suponer que las p+ q variables son centradas, es decir:

$$\sum_{i=1}^{n} p_i x_{ij} = 0 \qquad \text{con } j = 1, \ldots, p$$

$$\sum_{i=1}^{n} p_i y_{ik} = 0 \qquad \text{con } k = 1, \ldots, q$$

Con lo anterior se garantiza que el centro de gravedad de la nube de individuos $\{\overline{x}_i\}_{i \in I} \in R^p$ es el origen, e igualmente sucede con $\{\overline{y}_i\}_{i \in I} \in R^q$

II.4.3 Subespacios vectoriales

Por las combinaciones lineales de $\{x_j\}$ y $\{y_k\}$ se les asocia un subespacio vectorial de R^n: Vx y Vy respectivamente, a los conjuntos de vectores $\{x_j\}$ y $\{y_k\}$

II.4.4 Potencial de Previsión

A Vx se le llamará el potencial de previsión de $\{x_j\}$ donde:

$Vx = \{\lambda \in R^n \mid \lambda = x\bar{a};\ a \in R^p\} \subset R^n$ con Dim (Vx) < = p (Vx es un subespacio que limita la dispersión)

Y a Vy se le llamará el potencial de previsión de $\{y_k\}$ donde:

$Vy = \{\eta \in R^n \mid \eta = y\bar{b};\ b \in R^q\} \subset R^n$ con Dim (Vy) < = q

En este trabajo se supondrá que:

Rango (X) = p	Rango (Y) = q
Dim (Vx) = p	Dim (Vy) = q

Como anteriormente se supuso que las variables serían centradas, esto implica que λ y η son combinaciones lineales de variables centradas, sean también variables centradas.

Ahora, es interesante saber, qué tanto se parecen los subespacios Vx y Vy; si éstos son iguales, los conjuntos de variables son equivalentes. Como la igualdad es poco usual de darse, se buscarán las semejanzas.

Primero se obtendrán las combinaciones lineales de $\{x_j\}$ que pueden expresarse como combinaciones lineales de $\{y_k\}$ es decir, cuáles a_j^* y b_j^* cumplen:

$$\bar{\lambda} = x\overline{a}^* = y\overline{b}^* = \bar{\eta}$$

y que las parejas de vectores $\bar{\lambda} \in Vx$ y $\bar{\eta} \in Vy$ formen un ángulo mínimo:

$$\cos\theta = \frac{\operatorname{cov}(\bar{\lambda}, \bar{\eta})}{\sqrt{\operatorname{var}(\bar{\lambda})}\ \sqrt{\operatorname{var}(\bar{\eta})}} = \operatorname{corr}(\bar{\lambda}, \bar{\eta}) \rightarrow \overline{0}$$

Otro aspecto que se busca es la ortogonalidad en la bases $\{\bar{\lambda}_i\}$ y $\{\bar{\eta}_i\}$ de Vx y Vy respectivamente; por lo que se tendrá:

$$\bar{\lambda}_i \mid_{\Omega_p} \bar{\lambda}_j \quad y \quad \bar{\eta}_i \mid_{\Omega_p} \bar{\eta}_j \quad \text{con } i \neq j$$

De ahora en adelante las variables que se han manejado y otras que se usarán tendrán su propio nombre dentro del análisis canónico:

$\bar{\lambda}$ y $\bar{\eta}$: variables canónicas

$\overline{a} \in R^p$ y $\overline{b} \in R^q$: factores canónicos

$\mu = \sigma(\bar{\lambda}, \bar{\eta})$: correlación canónica

Entrando de lleno al problema, se van a buscar $\bar{\lambda} \in Vx$ y $\bar{\eta} \in Vy$ que no sean las triviales, que cumplan la condición de que su distancia sea mínima.

Si se supone que λ es conocido, el vector más cercano de Vy a éste, es su propia proyección Ω_p-ortogonal en Vy. Se toma a A_1 como una aplicación lineal que realiza la proyección, entonces $A_1\lambda$ es la imágen buscada; se toma ahora a $\overline{\eta} \in A_1\overline{\lambda}$, esto es un vector colineal a A_1 y se tiene que máx $\cos^2(\lambda, A_1\lambda) =$ máx $\cos^2(\lambda, \eta)$.

Como $A_1\lambda$ es colineal a η entonces $A_1\lambda = \alpha\eta$ y suponiendo:

$$\lambda = \eta = 1$$

se tiene que:

$$\alpha = A_1\lambda = \overline{\lambda} \cos(\overline{\lambda}, \overline{\eta})$$

y entonces:

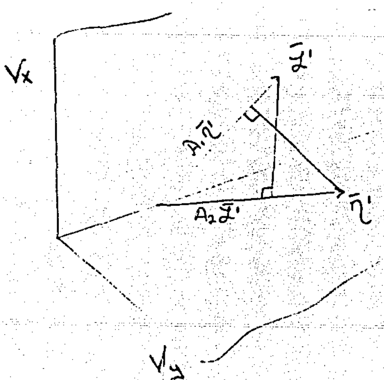
$$A_1\overline{\lambda} = \cos(\overline{\lambda}, \overline{\eta})\, \overline{\eta}$$

Análogamente, lo anterior puede hacerse con A_2 que sería la aplicación lineal que proyectaría η en Vx, con lo que tendría:

$$A_2\overline{\eta} = \cos(\overline{\lambda}, \overline{\eta})\, \overline{\lambda}$$

Así se tiene

$$(1) \qquad \begin{cases} A_1\overline{\lambda} = \cos(\overline{\lambda}, \overline{\eta})\, \overline{\eta} \\ A_2\overline{\eta} = \cos(\overline{\lambda}, \overline{\eta})\, \overline{\lambda} \end{cases}$$



Multiplicando las igualdades anteriores por la izquierda por A_2 y A_1 respectivamente se obtiene:

$$\begin{cases} A_2A_1\lambda = \cos(\overline{\lambda},\overline{\eta})\,A_2\overline{\eta} = \cos^2(\overline{\lambda},\overline{\eta})\,\overline{\lambda} \\ A_1A_2\overline{\eta} = \cos(\overline{\lambda},\overline{\eta})\,A_1\overline{\lambda} = \cos^2(\overline{\lambda},\overline{\eta})\,\overline{\eta} \end{cases}$$

Sea $\mu = \sigma^2 = \cos^2(\overline{\lambda},\overline{\eta})$ con $0 \leq \mu \leq 1$

Substituyendo en el sistema de ecuaciones (1):

$$\begin{cases} A_2A_1\overline{\lambda} = \mu\overline{\lambda} \\ A_1A_2\overline{\eta} = \mu\overline{\eta} \end{cases}$$

De donde $\overline{\lambda}$ y $\overline{\eta}$ son los vectores propios de los operadores A_1A_2 y A_2A_1 respectivamente, ambos asociados al valor propio λ

II.4.5 Vectores propios

$\overline{\lambda}$ y $\overline{\eta}$ son los vectores propios de A_1A_2 y A_2A_1, las siguientes parejas de variables canónicas resultan ser los vectores propios de A_1A_2 y A_2A_1 correspondientes a los valores propios ordenados en forma decreciente. Se pueden encontrar tantas variables canónicas como min(p,q).

Ahora para encontrar los factores canónicos se toma:

$$\begin{cases} A_2A_1\overline{\lambda} = \mu\overline{\lambda} \\ A_1A_2\overline{\eta} = \mu\overline{\eta} \end{cases}$$

substituyendo λ, η por xa y xb respectivamente se tiene:

$$A_1 A_2 xa = \mu xa$$

$$A_2 A_1 yb = \mu yb$$

substituyendo A_1 y A_2 por las igualdades:

$$\left(X\left(X^t \Omega_p X\right)^{-1} X^t \Omega_p\right)\left(Y\left(Y^t \Omega_p Y\right)^{-1} Y^t \Omega_p\right) X\bar{a} = \mu X\bar{a}$$

$$\left(Y\left(Y^t \Omega_p Y\right)^{-1} Y^t \Omega_p\right)\left(X\left(X^t \Omega_p X\right)^{-1} X^t \Omega_p\right) Y\bar{b} = \mu Y\bar{b}$$

de donde se define:

$$Vxx = X^t \Omega_p \; X$$

$$Vyy = Y^t \Omega_p \; Y$$

que son las matrices de varianzas y covarianzas de $\{x_i\}$ y $\{y_i\}$ respectivamente, y $Vxy = X^t \Omega_p Y = V^t yx$ es la matriz de las covarianzas entre $\{x_i\}$ y $\{y_i\}$.

Volviendo a las ecuaciones originales

$$XV^{-1}_{xx} V_{xy} V^{-1}_{yy} V_{yx}\, a = \mu xa$$

$$YV^{-1}_{yy} V_{yx} V^{-1}_{xx} V_{xy}\, b = \mu xb$$

y como se había supuesto que:

$$\text{rango } (x) = p$$

$$\text{rango } (y) = q$$

entonces:

$$V^{-1}_{xx} V_{xy} V^{-1}_{yy} V_{yx} a_i = \mu_i a_i$$

$$V^{-1}_{yy} V_{yx} V^{-1}_{xx} V_{xy} b_i = \mu_i b_i$$

que implica que a_i y b_i sean los vectores propios de $V^{-1}xx\ Vxy\ V^{-1}yy\ Vyx$ y $V^{-1}yy\ Vyx\ V^{-1}xx\ Vxy$ respectivamente derivados del mismo valor propio μ_i

Se interpretará a continuación los resultados:

se han construído bases Ω_p ortogonales de Vx y Vy, esto es:

$\{ \mu_i \mid i = 1, ..., p\}$ es base Ω_p ortogonal de Vx

$\{ \eta_i \mid i = 1, ..., q\}$ es base Ω_p ortogonal de Vy

Se toman los ejes μ^1 y μ^2 de Vx que proporcionan un plano, tomando las variables y^j con j = 1, ..., q, si se proyectan en el plano que se acaba de tomar en Vx se obtiene el siguiente vector: $y^j = A_1 y^j$ Vx cuyas coordenadas son:

$$\left(\overline{y}^{\,j}\, \Omega_p\, \mu^1 ,\ \overline{y}^{\,j}\, \Omega_p\, \mu^2 \right)$$

y por la simetría de A_1, resulta que:

$$\overline{y}^{\,j}\, \Omega_p\, \overline{\mu}^{\,l} = \overline{y}^{\,j} A^t_1\, \Omega_p\, \overline{\mu}^{\,l} = \overline{y}^{\,j}\, \Omega_p\, A_1\, \overline{\mu}^{\,l}$$

pero como $A_1\, \mu_i = \mu_i$ entonces:

$$\overline{y}^{\,j}\, \Omega\, \overline{\mu}^{\,l} = \overline{y}^{\,j}\, \Omega_p\, \overline{\mu}^{\,l}$$

y la variable $\{y_j\}$ es representada en el plano (μ^1, μ^2) por las coordenadas:

$$\left(\overline{y}^{\prime j}\,\Omega_p\,\mu^1\,,\ \overline{y}^{\prime j}\,\Omega_p\,\mu^2\right)$$

Tomando las variables más pequeñas, de tal manera que la varianza sea 1, $(S^2 = 1)$ entonces:

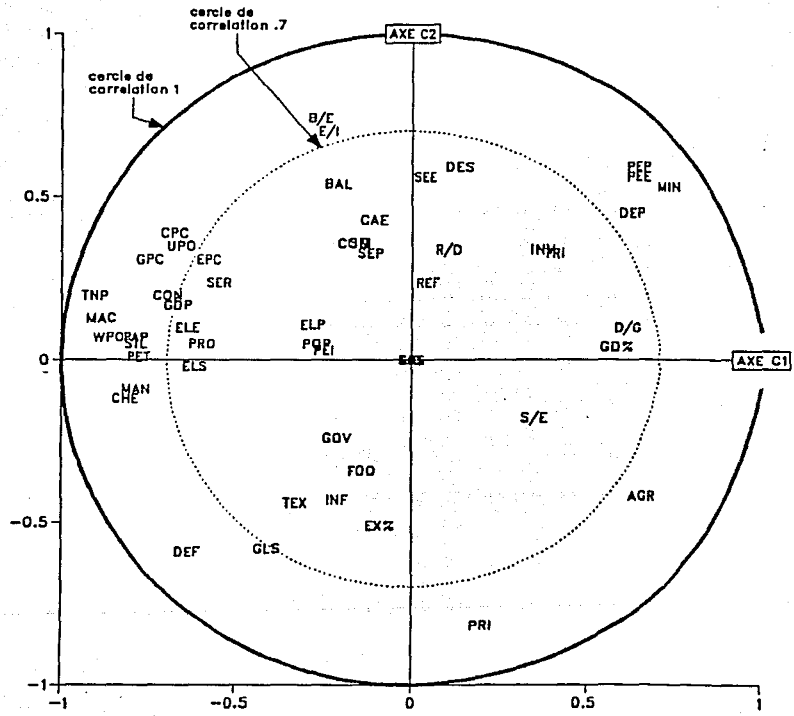
$$\overline{x}^{\prime j}\,\Omega_p\,\overline{\mu}^k = \cos\left(\overline{\mu}^k, \overline{x}^j\right) = \mathrm{corr}\left(\overline{\mu}^k, \overline{x}^j\right)$$
$$\overline{y}^{\prime j}\,\Omega_p\,\overline{\mu}^k = \cos\left(\overline{\mu}^k, \overline{y}^j\right) = \mathrm{corr}\left(\overline{\mu}^k, \overline{y}^j\right)$$

y como

$$L = \|\overline{x}^j\|^2_{\Omega_p} = \sum_{k=1}^{p}\left(\overline{x}^{\prime j}\,\Omega_p\,\overline{\mu}^k\right)^2 = \sum_{k=1}^{p}\cos^2\left(\overline{\mu}^k, \overline{x}^j\right) =$$
$$\|\overline{y}^j\|^2_{\Omega_p} = \sum_{k=1}^{p}\cos^2\left(\overline{\mu}^k, \overline{y}^j\right) \Rightarrow \cos^2\left(\overline{\mu}^1, \overline{x}^j\right) + \cos^2\left(\overline{\mu}^2, \overline{x}^j\right) \leq 1$$

y es así que cada variable queda inscrita en un círculo con centro en el origen y radio 1. Por las características de definición del círculo, se le llama círculo de correlaciones. La utilidad del círculo es que, ayuda en las semejanzas que puedan existir entre las variables por medio del ángulo que forman*.

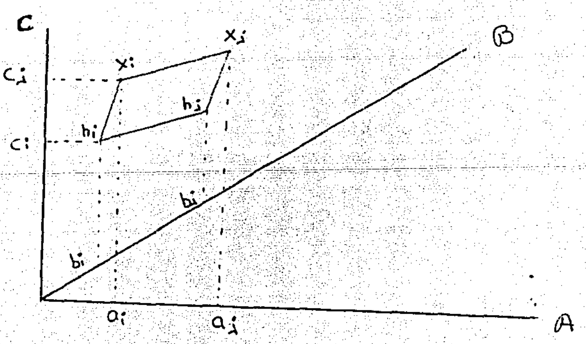
* Mc. Gregor A., Frederic, *Analyse des structures économico-industrielles et de ...*, p. 28.


cercle de correlation .7
cercle de correlation 1
AXE C2
AXE C1
B/E
E/I
BAL
SEE
DES
PEP
PEE
MIN
DEP
CAE
CPC
UPO
GPC
EPC
CSR
SEP
R/D
INV
PRI
SER
REF
TNP
CON
GDP
MAC
ELE
ELP
D/G
WPO
PAP
SIL
PRO
POP
PEI
GD%
PET
ELS
MAN
CHE
S/E
GOV
FOO
TEX
INF
AGR
EX%
DEF
GLS
PRI
1
0.5
0
−0.5
−1
0.5

II.5 Análisis de Componentes Principales

A fin de comprender mejor la idea del análisis de componentes principales, veámoslo con tres variables x1,x2 y x3, observemos las distancias entre los diferentes individuos gráficamente, esto es algo difícil pues no se aprecia muy bien la distancia entre todos ellos, es por ello que se toma una línea "Y" y se proyectan en ella los puntos. Por analogía se puede reducir un conjunto de puntos de R^3 a R^2 para poder analizar visualmente las relaciones entre ellos, claramente las variables no están siempre bien correlacionadas, pues cada vector tiene su propia métrica.

El método consiste, en encontrar los ejes de proyección del plano que maximicen las distancias originales. Este problema puede resolverse buscando los ejes donde: $d^2(x_i, x_j)$ sea máxima.



La distancia $d^2(h_i, h_j)$ es la proyección ortogonal sobre el plano H de la distancia $d^2(x_i, x_j)$, es claro que $d^2(h_i, h_j) <= d^2(x_i, x_j)$. Por otro lado $d(a_i,a_j)$ es la proyección sobre el eje A y $d^2(b_i,b_j)$ es la proyección sobre el eje B.

Para encontrar el eje de proyección A, se maximiza $d^2(a_i, a_j)$, y de la misma forma, el resultado de maximización de $d^2(b_i, b_j)$ es el eje B; de esta forma con dos ejes perpendiculares se puede formar el plano H. Si existe una pérdida considerable de información, se puede calcular un tercer eje de proyección perpendicular al plano que puede ser encontrado de la misma forma. El error que se sigue es el de formar n ejes que formen otro sistema de coordenadas en donde se tenga toda la información de la nube original pero que nuevamente es difícil de ver.

II.5.1 Centro de gravedad

Para simplificar el proceso se translada el origen al centro de gravedad de la nube, es decir:

$$g = 0 \qquad R^3$$

Se trabaja sin unidades, pues, el tenerlas implica decidir bajo qué unidades referenciar las medidas.

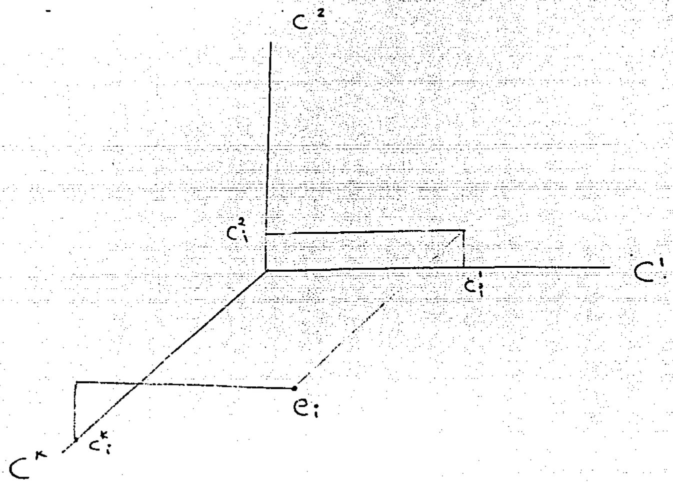
De la sección anterior se tiene:

$$\cos \theta_{xy} = \frac{S_{xy}}{S_x S_y} \qquad \frac{\text{varianza}}{\text{varianza de x} \quad \text{varianza de y}}$$

De esta forma la distancia entre las variables definida por el coseno entre ellas es igual a su coeficiente de correlación.

Ahora para encontrar los ejes en donde las distancias de los individuos sean máximas, se toma de nuevo el espacio definido por las variables.

Un punto e_i tiene en el origen las siguientes coordenadas: $x^1{}_i$, $x^2{}_i$, $x^3{}_i$.



la proyección de e_i sobre los ejes principales (llámense éstos C^1, C^2, C^3) tienen como coordenadas $c^1{}_i$, $c^2{}_i$, $c^3{}_i$. La proyección de todos los puntos e_i sobre un eje principal cualquiera C^k compone un vector.

Una componente principal C^k es entonces una simple combinación lineal de variables del origen:

$$C^k = U^k{}_1 x^1 + U^k{}_2 x^2 + U^k{}_3 x^3$$

sea $c = XU$

Dado que los datos están centrados, la varianza de un nuevo vector C_j puede determinarse de la forma siguiente:

$$S^2_{cj} = < c_j, c_j > = U^t X^t X U = U^t V U$$

donde V es la matriz de varianza-covarianza de las variables del origen y u es un vector unitario. Se buscará entonces, el vector u que define el eje C^j que maximiza la varianza S^2_c

Este problema puede resolverse buscando el valor máximo del radio

$$U^t V U \,/\, U^t U$$

$$\frac{d}{du}\left(\frac{U^t V U}{U^t U}\right) = 0 \quad => \quad V_U = \left(U^t V U \right) U = \lambda_u$$

El problema de buscar los ejes que maximizan la varianza puede entonces ser resuelto con los valores propios asociados a la matriz de varianza-covarianza de los datos originales.

II.5.2 Valores propios y vectores propios

Dado que esta matriz es simétrica, sus vectores propios, asociados son ortogonales*. Es así como los vectores propios constituyen un nuevo sistema de ejes ortogonales no correlacionados entre sí. La varianza máxima es igual al más grande valor propio sobre el eje definido por su vector propio asociado. Si los valores propios se ordenan en forma decreciente se pueden fácilmente encontrar el 2o., 3er, etc. ejes principales. La matriz de varianza Vc (matriz de pesos) del nuevo sistema es entonces la matriz diagonal compuesta por los valores propios asociados a la matriz de varianza de los datos originales.

* Burden, R. *Análisis Numérico*, p. 491.

III. ANALISIS FACTORIAL DE CORRESPONDENCIAS

En este capítulo se va a presentar el programa que realiza el análisis de correspondencias, y mediante el seguimiento de un ejemplo, se va a describir tanto el programa como el método que emplea.

El ejemplo que se va a seguir es el siguiente: FUMADORES DE UNA EMPRESA CONTRA LA CATEGORIA DE SU PUESTO. El análisis se va a efectuar con una tabla de contingencia de 5x6, con datos de 193 personas de una empresa*. Los renglones serán las categorías de empleo y las columnas serán las categorías de fumadores y los dos últimos datos adicionales o suplementarios se refieren a los que consumen alcohol.

	Consumo de tabaco				Consumo de alcohol	
	(1) No	(2) Bajo	(3) Medio	(4) Alto	(1) No	(2) Si
(1) Gerencia	4	2	3	2	0	11
(2) Subgerentes	4	3	7	4	1	17
(3) Empleados maduros	25	10	12	4	5	46
(4) Empleados jóvenes	18	24	33	13	10	78
(5) Secretarias	10	6	7	2	7	18
Promedio	42%	29%	20%	9%		

* Grenacre, M. J., *Theory and applications of ...,* p.72

III.1 Esquema general

El esquema general del proceso se muestra a continuación*

- Proceso general
  - Programa ADC
    - LECTUR
    - VPROP
      - TRIDI
    - VPOUR
    - FACTOR
    - SORTIE
  - Programa ADD
    - COORD
      - MAXBE
        - TRI
        - CONSER
    - WRT

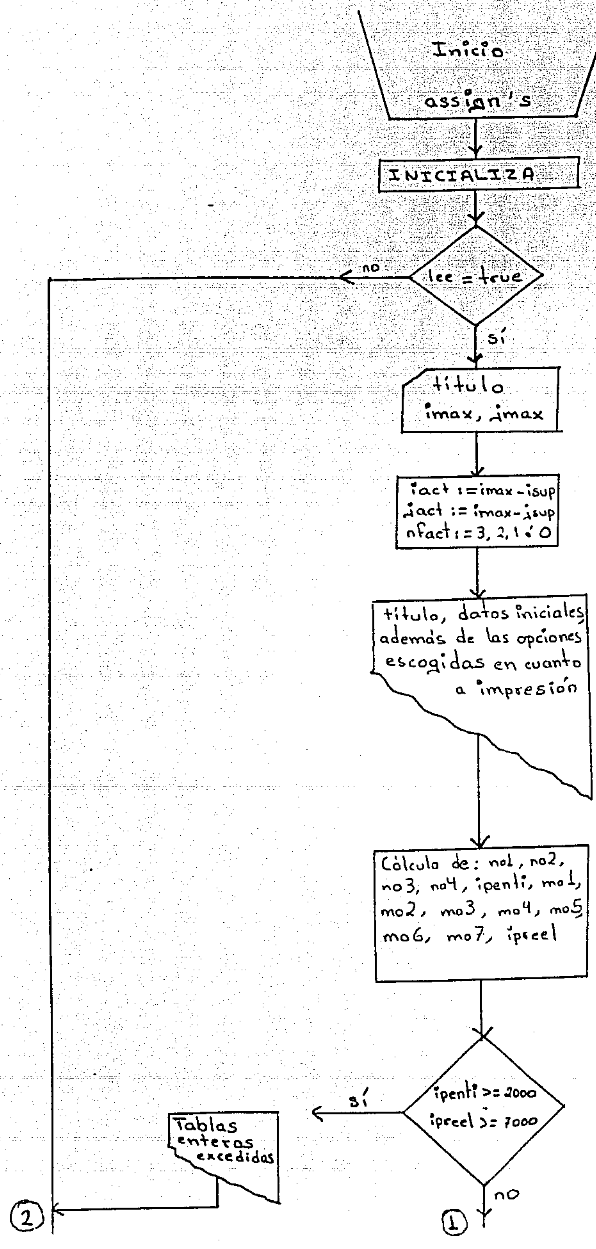
La ejecución de los programas ADC y ADD requiere de los datos necesarios para el análisis; éstos son almacenados en los archivos DATOS1 y DATOS2.

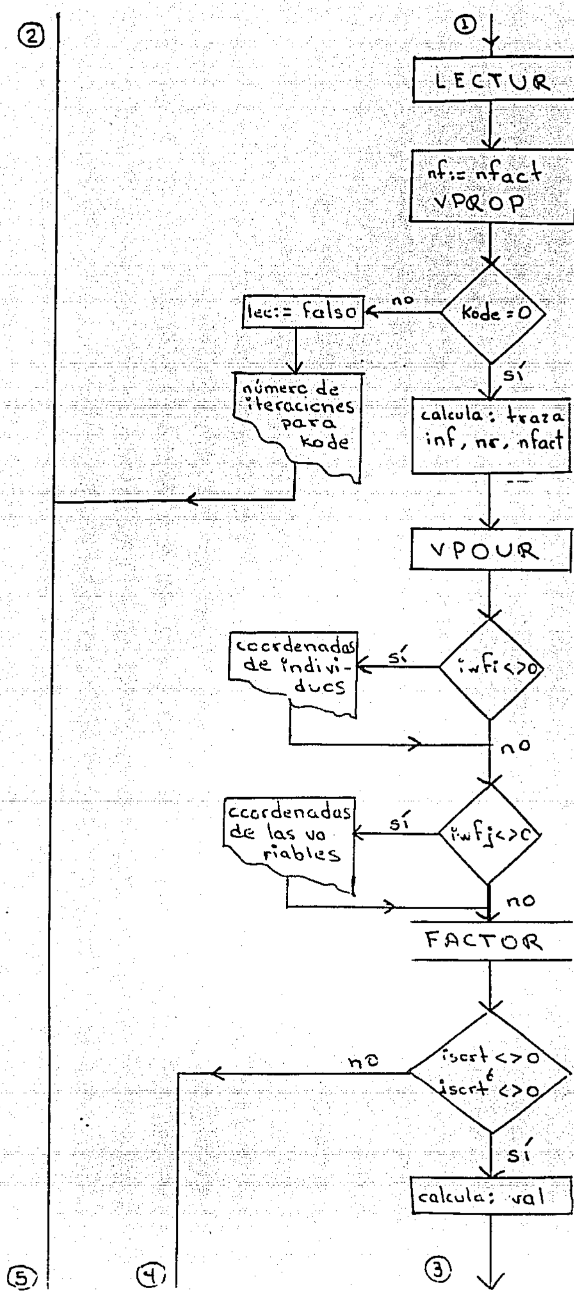
Se utilizan además otros archivos: TRANS que sirve de enlace para ADC y ADD; DATOS3, DATOS4 y DATOS5 que se utilizan para almacenar datos temporales; CORIN y CORVAR son archivos que guardan las coordenadas de los individuos y de las variables respectivamente.

El diagrama de flujo de ADC es el siguiente:

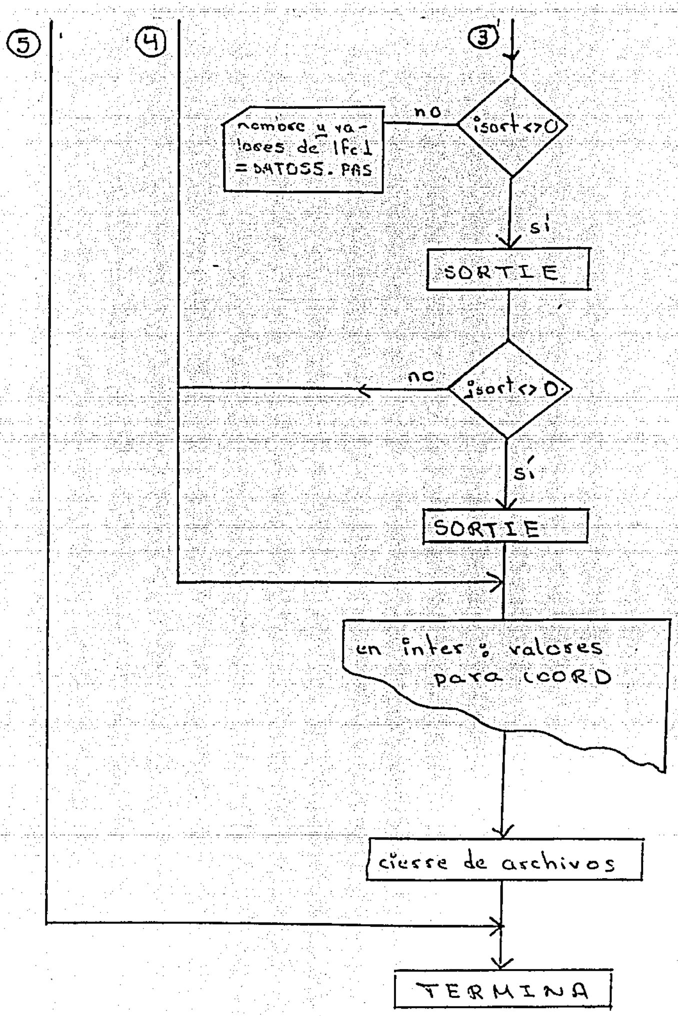
* los diagramas de flujo más importantes, se mostrarán conforme se presenten los módulos que componen a los programas .

Diagrama de flujo de ADC





②
①
LECTUR
nf:= nfact
VPROP
Kode = 0
no
lec:= falso
sí
número de iteraciones para kode
calcula: traza inf, nr, nfact
VPOUR
iwfi<>0
sí
coordenadas de individuos
no
iwfj<>0
sí
coordenadas de las variables
no
FACTOR
iscrt <>0 ó iscrt <>0
no
sí
calcula: val
⑤
④
③


5
4
3
isort<>0
no
nombre y valores de lfc1 = DATOS5. PAS
si
SORTIE
isort<>0
no
sí
SORTIE
en inter: valores para ICORD
cierre de archivos
TERMINA

III.2 Inicialización de datos

El programa comienza con la definición de los archivos de entrada, de salida, los de interface y los de almacenamiento de datos parciales. A continuación se inicializan las variables (que pueden considerarse como constantes, y no están en la sección de *const* de Pascal porque alguna subrutina de los programas con ciertas características puede modificarlas, para registrar los datos los lee del archivo DATOS2).

Entre los datos que se leen está el número de individuos suplementarios [0]* (filas) y de variables suplementarias [2] (columnas), el número de factores con respecto a los cuales se hará el análisis y que por supuesto se tomarán para formar las gráficas [2] (ejes) y otros que son indicaciones de si se quieren gráficas o no [SI], si se quieren almacenar datos de los individuos o de las variables [NO, AMBOS] y si estos se quieren o no imprimir [SI AMBOS]. Se calculan algunas variables con los datos anteriores y se continua con la lectura del título del análisis a efectuar [FUMADORES DE UNA EMPRESA CONTRA LA CATEGORIA DE SU PUESTO], y el número máximo de individuos que es la suma de los activos y los suplementarios [5].

Se activa otra rutina de lectura la cual precisamente leerá en primer lugar los nombres de las variables activas [NOFU = no fuman, POCO = fuman poco, MEDI = mediano consumo de tabaco, MUCH = fuman mucho, NOBE = no beben alcohol, SIBE = si beben alcohol], las variables suplementarias –las últimas dos– van al final. Es muy importante conocer que el límite de variables es 90**, después se leen los comandos gráficos, esto es, cómo se quieren las gráficas, se solicitarán dos: una con elementos activos solamente y la otra con activos y

* los datos encerrados en paréntesis cuadrados indican que son los datos que a la computadora se le darán para que corra el programa, el significado de cada variable está descrito en el capítulo IV – MANUAL DE USO–

** PASCAL para PC versión 3.0 tiene limitado el número de *array's* y la dimensión de ellos.

suplementarios. Y asi como se tiene un número máximo de variables y de individuos, también lo hay para las gráficas: 16.

III.3 Proceso

Al terminar la lectura de los comandos gráficos, comienza a leer la matriz de datos (X_{nxp}) iniciando con el nombre del individuo y la línea de información de éste con respecto a cada una de las variables, en otras palabras; lee por filas la matriz; como en el caso de las variables, es conveniente conocer que el número máximo de individuos es 90. Dado el orden de lectura, primero tomará a los individuos activos y después a los suplementarios [jeav = jefes de edad avanzada, jejo = jefes jóvenes, emav = empleados de edad avanzada, emjo = empleados jóvenes, secr = secretarios (as)]

[illegible]	[illegible]	[illegible]	[illegible]	MEDI	MUCH	[illegible]	SIEE
[illegible]	?	41	45	43	25	23	170
	?	4	2	3	2	0	11
	?	1	3	7	4	1	17
	?	25	10	12	4	5	46
	?	12	24	33	13	10	79
	?	10	[illegible]	7	3	7	19

III.4 Cálculo de valores y vectores propios

En el caso de que se quieran imprimir la matriz de datos, el siguiente paso es dicha impresión, si no, termina la rutina y el programa principal ejecuta el siguiente procedimiento que es el cálculo de vectores y valores propios de X_{nxp}.

Para obtenerlos, se transforma la matriz de datos en una matriz triangular inferior, se multiplica por su transpuesta y así se obtiene una matriz simétrica que luego se tridiagonaliza por medio de la subrutina TRIDI basado en el método de Householder y con ésta aplicar el método QL para finalmente encontrar los valores y los vectores propios de la matriz*.

Después de obtener los valores propios y los vectores propios, calcula la traza de la matriz, enseguida ejecuta la rutina VPOUR la cual imprime los valores propios, el porcentaje, el porcentaje acumulado y un histograma de los valores propios, calculados de la siguiente manera:

prc = porcentaje: 100 * valor propio / traza

pcum = porcentaje acumulado: pcum + prc

histograma: P (1 + 70 * valor propio / valorp[2])

```
EDICION DE VALORES PROPIOS
--------------------------

EL PRIMER VALOR PROPIO (PARASITO) ES ELIMINADO   1,0000000E+00

SUMA DE LOS VALORES PROPIOS ACTIVOS          9,5169546E-02

----------------------------------------------------------------------------------------------------
?    ? VALOR PROPIO ? PORCENTAJE ? PORCENTAJE ? HISTOGRAMA DE LOS PRIMEROS VALORES PROPIOS
?    ?              ?            ? ACUMULADO ?
----------------------------------------------------------------------------------------------------
?  1 ? 7.4772107E-02 [illegible] 7,85E+01 ? 7,85E+01 ? **************************************************************
?  2 ? 1,9917190E-02 [illegible] 2,09E+01 ? 9,95E+01 ? *********
?  3 ? 4,1357407E-04 [illegible] 4,95E-01 ? 1,00E+02 ? *
```

* Burden, Richard L. *Análisis Numérico*, p. 512

A continuación pregunta si se conservarán las coordenadas de individuos y de las variables, las respuestas están en el archivo DATOS2.PAS, si la variables iwfi es igual a 1 los guardará, si la variable iwfj es 1 guardará las coordenadas de las variables, si son 0 no lo hará. En el caso que se está siguiendo sí las guardará:

de los individuos son:

de las variables son:

Después ejecuta el procedimiento FACTOR el cual comienza a manipular todos los datos para calcular las gráficas y las proyecciones (en el 2º eje).

III.5 Impresión de resultados

El siguiente paso es imprimir las tablas de los individuos y de las variables, siempre y cuando el valor de isort y de jsort sean diferentes de cero, porque de lo contrario no las imprimirá (para este ejemplo sí se imprimirán). Si se ha pedido la impresión de alguna tabla se ejecuta el subprograma SORTIE con los parámetros adecuados a lo que se solicita, las siguientes son las tablas de este ejemplo:

? I1 ?	CAL	FEED	INE?	1=F	COR	CTR?	2=F	COR	CTR?
1?JEAU?	992	570	31?	-45	92	37	194	906	214?
[illegible]	?91	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	243	[illegible]	[illegible]
[illegible]1000	2440	[illegible]	[illegible]	[illegible]	[illegible]	11	1	3?	
[illegible]1000	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-57	[illegible]	[illegible]	
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-7?	1?3	[illegible]
? ?		193	1000?			1000?			1000?

[illegible]	CAL	FEED	INE?	1=F	COR	CTR?	2=F	COR	CTR?
[illegible]1000	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	7?		1	29?
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-140	657	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-4	1	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
? ?		1??	1000?			1000?			1000?

[illegible]	CAL	FEED	INE?	1=F	COR	CTR?	2=F	COR	CTR?
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
		[illegible]	[illegible]			[illegible]			17???

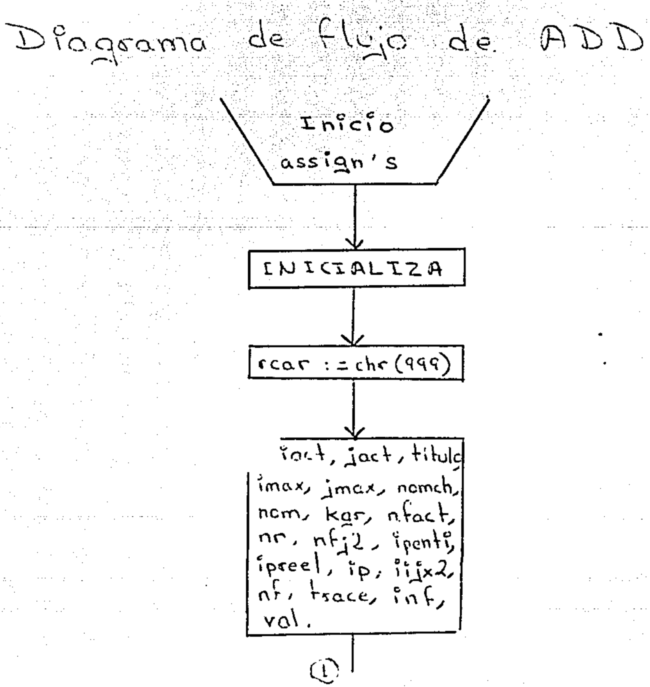
FALLA DE ORIGEN

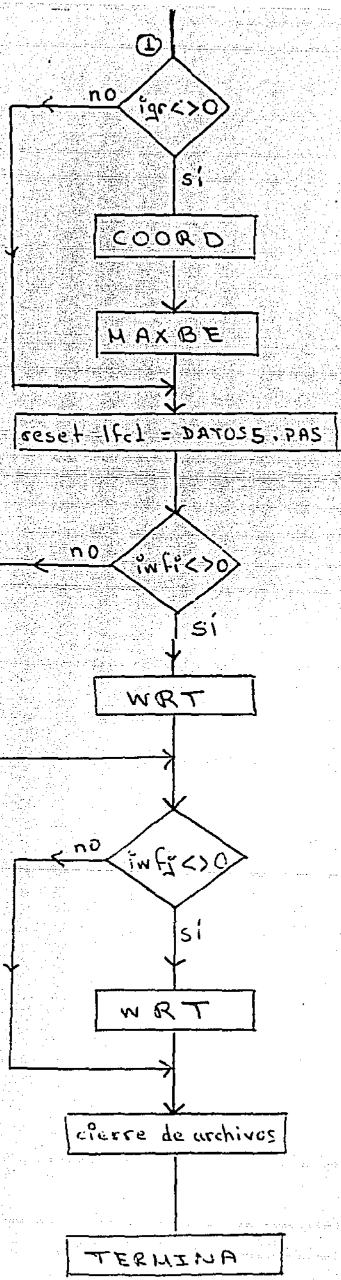
La última parte del programa principal de ADC.PAS es la que se encarga de escribir los datos necesarios para la impresión de las gráficas, que realiza el programa ADD.PAS.

Es oportuno aclarar que la salida de ADC.PAS es almacenado en el archivo nombrado: SALIDA.PAS a menos que se quiera cambiar directamente en el programa las instrucciones pertinentes para que la salida sea por pantalla o por la impresora.

III.6 Impresión de gráficas

El programa ADD.PAS es el encargado de imprimir las gráficas con las características que se le dieron en DATOS1.PAS. Esta impresión es ejecutada en el archivo SALIDA.TXT. El diagrama de flujo de ADD es el siguiente:




①
no
igr<>0
sí
COORD
MAXBE
reset lfc1 = DATOS5.PAS
no
iwfi<>0
sí
WRT
no
iwfj<>0
sí
WRT
cierre de archivos
TERMINA

El programa principal comienza con la inicialización de las variables, esto es, la lectura del archivo que sirve de enlace entre ambos programas -ADC.PAS envia algunos valores de variables a ADD.PAS por medio de este archivo- y si igr es diferente de cero entonces activa el subprograma COORD.

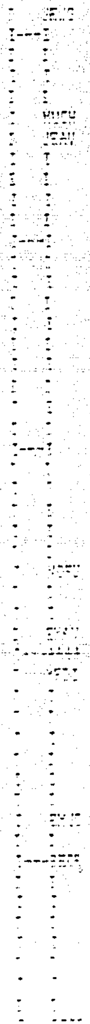
COORD lee los datos que proporcionan las características de la primera gráfica y comienza a separar los elementos según esas características y ejecuta el subprograma MAXBE, el cual, con los datos que le proporciona como parámetros, comienza a imprimir los ejes y, en las coordenadas de los puntos los nombres de los elementos que se le piden.

Al finalizar el subprograma MAXBE, el programa principal sigue con el almacenamiento de las coordenadas de los individuos (si iwfi = 0) y de las variables (si iwfj = 0) activando el subprograma WRT el cual lee el nombre del individuo o variable y sus respectivas coordenadas de un archivo temporal donde las ha puesto, dejando en el archivo CORIN.PAS a los individuos, y en CORVAR.PAS a las variables. El programa principal termina cerrando todos los archivos.

Así se observa, que la salida está separada en dos archivos: datos, tablas, histograma y demás están en SALIDA.PAS (que ya han sido expuestos) y las gráficas y proyecciones generadas por ADD.PAS en SALIDA.TXT. Las gráficas para este ejemplo se muestran a continuación:

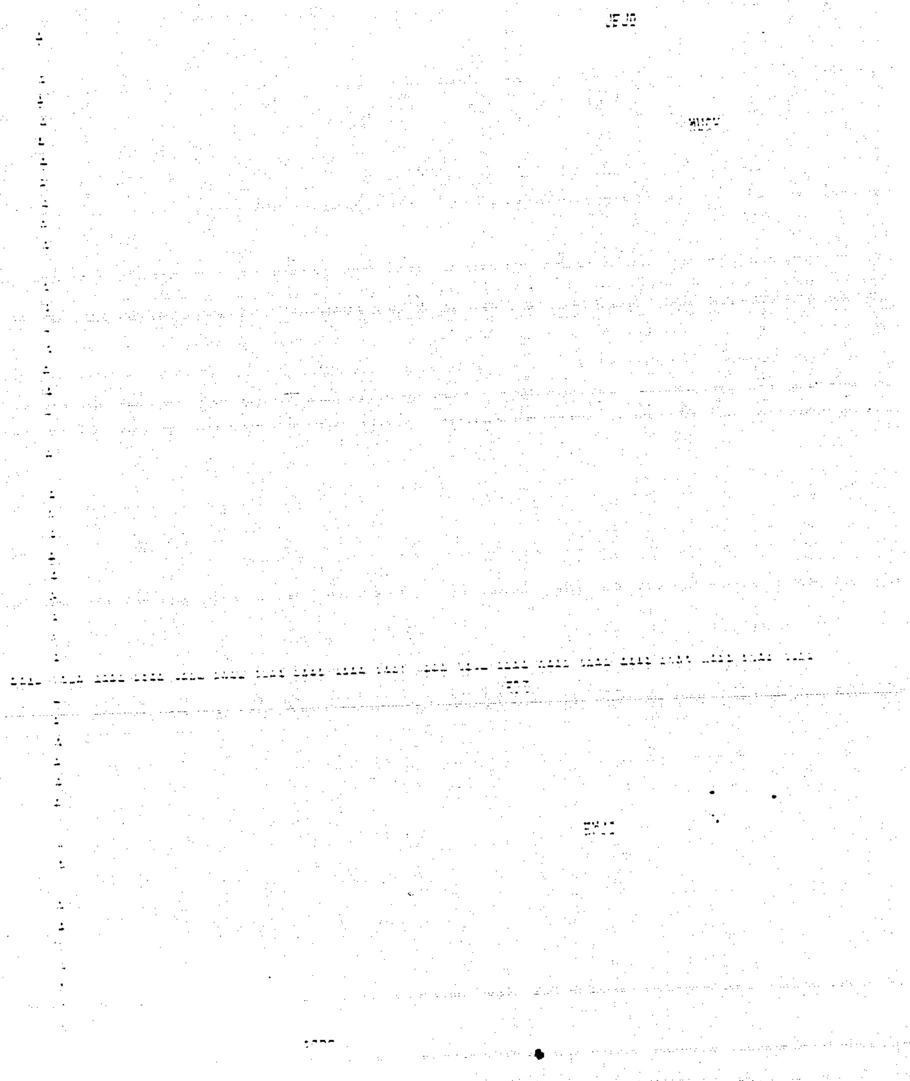
9 PUNTOS EN EL PLANO DETERMINADO POR LOS EJES 1 Y 2

VALORES EXTREMOS SOBRE EL EJE 2 -1,41064289E-01 2,43304574E-01
INTERVALO ENTRE DOS LINEAS 7,69704465E-03



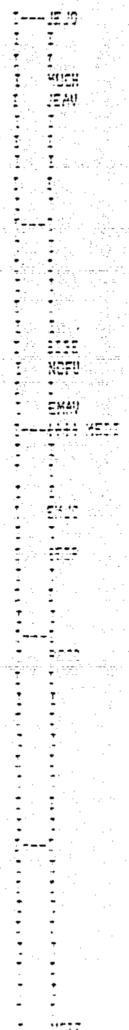
FUMADORES DE UNA EMPRESA CONTRA LA CATEGORIA DE SU PUESTO
EJES 1 Y 2

FALLA DE ORIGEN



11 PUNTOS EN EL PLANO DETERMINADO POR LOS EJES 1 Y 2

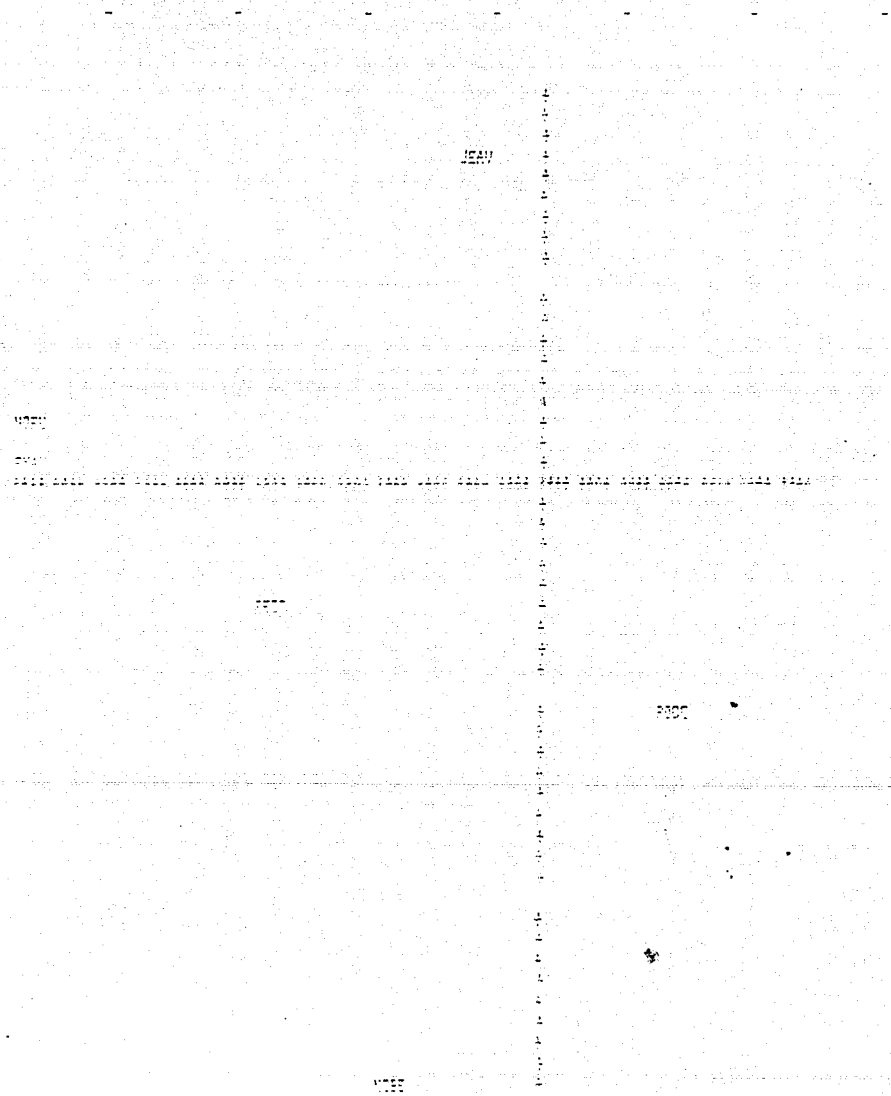
VALORES EXTREMOS SOBRE EL EJE 2 -3,61955979E-01 2,43304574E-01
INTERVALO ENTRE DOS LINEAS 1,21177316E-02



VALORES EXTREMOS SOBRE EL EJE [illegible]
INTERVALO ENTRE DOS COLUMNAS [illegible]

FALLA DE ORIGEN

FUMADORES DE UNA EMPRESA CONTRA LA CATEGORIA DE SU PUESTO
EJES 1 Y 2



SUMARIO

IV.1.- Introducción

El programa de Análisis Factorial de Correspondencias fue hecho por Jean Paul Benzécri (programa TABET) en lenguaje FORTRAN, y esta presente versión, está adaptada para correr en microcomputadoras PC o compatibles bajo el sistema operativo MSDOS versión 2.11 o mayor con al menos 512 Kbytes de RAM, el programa está escrito en TurboPascal versión III.

Este programa hace un análisis de correspondencias de una matriz de datos con un número de líneas igual al valor del parámetro imax, donde éstas representan a los individuos y existen tantas columnas en la matriz como lo indica el parámetro jmax, las columnas corresponden a las variables; se tiene además la posibilidad de dar individuos suplementarios incluídos en los imax individuos (el número de éstos lo proporciona el parámetro isup) y/o variables suplemetarias dentro de las jmax variables (la cantidad de variables suplementarias está en el parámetro jsup).

En el 1º y 2º capítulos se muestra el formato en el que los parámetros y datos deben ser introducidos para el correcto funcionamiento de los programas que efectúan el análisis de correspondencias.

En el último capítulo se expone la forma en que los resultados son mostrados al usuario, esto es, la secuencia de las impresiones y lo que cada una de ellas representa del análisis.

Y en la última parte de este capítulo, se dan las variantes que puede haber en la salida, diferentes formas de gráficas y sus respectivas proyecciones, así como la descripción de otros archivos relacionados con los resultados.

IV.1.1 Algunas Indicaciones

El programa que efectúa el análisis de correspondencia está dividido en dos partes: en la primera se realizan todas las operaciones, como el cálculo de valores propios, la inercia, etc., y la segunda parte se encarga de la impresión de las proyecciones y las gráficas a que corresponden las primeras. Ambas partes son dos programas llamados: ADC.PAS y ADD.PAS respectivamente, los cuales para funcionar correctamente necesitan de los archivos DATOS1.PAS y DATOS2.PAS que les proporcionan los datos necesarios, además hay otros archivos que serán utilizados, como los de las salidas, uno de interfase para ADC.PAS y ADD.PAS y otros que se utilizan temporalmente; cada uno de todos estos archivos será tratado más adelante.

Se pueden tener varias parejas de archivos con los datos iniciales, pero solamente los que se llamen DATOS1.PAS y DATOS2.PAS serán aceptados por ADC.PAS y ADD.PAS, por lo que se tiene la opción de cambiar nombres a las diferentes parejas de archivos que se quieran tener como datos aceptados, o cambiar el nombre de DATOS1.PAS y DATOS2.PAS en los programas que los usan por los nombres de los archivos con los datos que se deseen. Este cambio se hace dentro de la sección de los assign's en los programas principales.

Para utilizar ADC.PAS y ADD.PAS se debe entrar al lenguaje TurboPascal y correr el pograma ADC.PAS, a continuación se corre el programa ADD.PAS. Debe efectuarse en este orden pues, ADD.PAS trabaja con los datos proporcionados por ADC.PAS.

Otra forma es: generar los programas objeto de ADC.PAS y ADD.PAS mediante TurboPascal, para correrlos se sigue el mismo orden pero ahora desde el sistema operativo se teclea solamente ADC y ADD.

IV.2 Archivo de parámetros

En el archivo de datos llamado DATOS1.PAS se tienen los parámetros que determinan que se impriman las tablas y gráficas. Estos parámetros deben ser conservados sin alteración pues se usan en el programa de cálculos estadísticos y en el programa que hace las proyecciones y las gráficas.

A continuación se presenta el orden en el que deben estar, y una pequeña explicación de los que significan o expresan.

TITRE : Título del análisis que se pretende hacer, consta de 80 caracteres alfanuméricos y aparecerá en cada representación gráfica

IMAX : Número total de individuos que ocupan las líneas de la matriz (incluye los suplementarios).

JMAX : Número total de variables que ocupan las columnas de la matriz (incluye las suplemetarias).

NOMCH : Es la lista de los nombres (cuatro caracteres cada uno) de las jmax variables; si hay variables suplementarias los nombres de éstas van al final de la lista.

KGR : Matriz de dimensión 5 x 16 que proporciona el número de gráficas deseadas (máximo 16) y en cada columna se describen las características de la gráfica correspondiente (la letra X significa el número de la gráfica):

KGR[1,X] = Número del factor puesto sobre el eje horizontal.

KGR[2,X] = Número del factor puesto sobre el eje vertical.

KGR[3,X] = 1 Para la representación de los individuos activos.

2 Para la representación de los individuos suplementarios.

3 Para todos los individuos (activos + suplementarios).

0 Para no representar estos individuos.

KGR[4,X] = 1 Para representar variables activas.

2 Para representar variables suplementa- rias.

3 Para representar todas las variables (activas + suplementarias).

0 Para no representar estas variables.

KGR[5,X] = 0

Con la información anterior, el archivo DATOS1.PAS debe estar como el siguiente ejemplo:

ANALISIS DE CORRESPONDENCIAS

2 3

NOM1 NOM2 NOM3

1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0

2 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0

0 3 0 0 0 0 0 0 0 0 0 0 0 0 0 0

1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

En los datos anteriores se dio:

— el nombre del análisis a efectuar

— hay dos individuos y tres variables

— los siguientes tres nombres corresponden a las tres variables

— se indica a continuación que se quieren dos gráficas: la primera con representación de individuos activos solamente; la segunda con una representación de todos los individuos activos y suplementarios y variables activas solamente. Los demás ceros deben darse e indican realmente que se quieren otras 14 gráficas, pero sin información. Si no se dan los ceros el programa no corre.

IV.3 Archivo de datos

En el archivo DATOS2.PAS se dan más parámetros que solamente serán utilizados en el programa ADC.PAS y además este archivo de datos tiene la información de los individuos, que también son ocupados solamente en ADC.PAS. A continuación se expone el orden en el se van a leer, y una breve descripción de su significado.

ISUP : Número de individuos suplementarios que corresponden a los imax individuos.

JSUP : Cantidad de variables suplementarias que corresponden a las jmax variables.

NFACT : Número de factores de extracción (máximo 7).

IGR : Si IGR = 1 La representación de gráficas es activada.

Si IGR= 0 Se hace caso omiso de la representación gráfica.

ISORT: Si ISORT= 1 Da la impresión de las estadísticas relacionadas a los individuos.

Si ISORT = 0 No imprime las estadísticas relativas a los individuos.

JSORT: Si JSORT= 1 Da la impresión de las estadísticas relacionadas a las variables

Si JSORT= 0 No se imprimen las estadísticas relacionadas a las variables

IWFI: SiIWFI = 1 Las coordenadas de los imax individuos sobre los nfact factores, son guardadas en el archivo LFCI.PAS.

IWFJ : Si IWFJ = 1 Las coordenadas de las jmax variables sobre los nfact factores son guardadas en el archivo LFCJ.PAS.

Los datos en este archivo deben estar en el orden en que se describieron, el siguiente ejemplo que es continuación del anterior lo muestra:

```
1 1 2 1 1 1 1 0
NS1
11 12 13
NS2
21 22 23
```

Lo anterior indica que:

— existe un individuo suplementario
— una variable suplementaria
— se sacarán dos factores de extracción
— como IGR es uno, indica que se imprimirán las gráficas
— se quiere la impresión de los resultados de los individuos
— se quiere la impresion de los resultados de las variables
— se guardarán las coordenadas de los imax individuos
— no se quiere guardar las coordenadas de las variables
— se proporciona el nombre del primer individuo
— después de cada nombre de individuo se dan los datos que corresponden a él, son tantos valores como jmax variables existan, en este caso jmax es igual a 3; en caso de que existan individuos suplementarios, éstos se colocan al final de la lista, como en caso presente donde existe uno.

IV.4 Salida

Los programas ADC.PAS y ADD.PAS tienen cada uno su propio archivo de salida, esto implica que existen dos archivos de salida; uno de nombre SALIDA.PAS que contiene: la matriz de datos, los valores propios, el histograma y demás tablas con resultados, es decir, todos los resultados de ADC.PAS; el otro archivo de salida llamado SALIDA.TXT contiene a todas las gráficas y sus respectivas proyecciones, todos estos resultados corresponden a los generados por ADD.PAS.

En el primer archivo las impresiones tienen el siguiente orden:

— Título del análisis que se hizo.

— El número de individuos y número de individuos suplementarios que son incluidos.

— El número de variables y número de variables suplementarias que son incluidas.

— La cantidad de factores de extracción que se quieren.

— Si hay otras impresiones, como los factores sobre el conjunto de individuos o variables, aquí se indica cuál de ellos, o ambas.

— Si hay variables suplementarias, se imprimen los nombres de éstas.

— La cifra de gráficas que se quieren. Si IGR = 0 no habrá salida de gráficas y aquí se mencionará esto si es el caso.

— Si IDAT = 1 se imprime toda la matriz de datos con los respectivos nombres de variables e individuos.

— Sigue la impresión de una tabla con los valores propios, los porcentajes de inercia asociados a cada eje factorial, los porcentajes acumulados y un histograma de los valores propios.

— Si IDAT 0 se imprimen los resultados relativos a todos los individuos, solo que, los suplementarios en una tabla a parte.

Las tablas se componen con los nombres de los individuos en las líneas, y las columnas con CAL, PESO, INE, nfact veces F, COR y CTR donde:

CAL = calidad de la representación sobre el espacio factorial.

PESO = peso de cada individuo (en 10 000-ésimas) de la población total.

INE = parte del punto de la inercia vp1 + ... + vpNFACT sobre la nube (cada punto cuenta para un número de veces igual a PESO), con vp = l-ésimo valor propio.

F = valor sobre el factor.

COR = calidad de representación sobre el factor.

CTR = parte del punto de la varianza del factor (cada punto cuenta un número de veces igual a PESO).

Si JSORT 0 se imprimen los resultados relativos a todas las variables, pero, las suplementarias están en otra tabla después de ésta.

Las tablas están dadas con las mismas características de las tablas para individuos, solamente que aquí salen los nombres de las variables y sus respectivos resultados.

IV.4.1 Otros archivos

Ya se vieron los archivos de datos y los de salida, ahora se describirán las características de los restantes y la importancia que tienen.

CONFIG.SYS.- Este archivo no está declarado dentro de ADC.PAS ni de ADD.PAS, pero es muy importante para éstos programas, pues son instrucciones para el sistema operativo; éstas modifican el número de archivos y buffers que son aceptados en ADC.PAS y ADD.PAS; es indispensable que se encuentre en el disquette o en un disco duro (lo que la máquina acepte primero y donde se encuentra el sistema operativo esto puede verificarse listando la totalidad de archivos. En caso de no tenerlo se puede crear con el copy command del sistema operativo y dar:

files= 10

buffers= 10

en caso de tenerlo pero con menos cantidad, solamente habrá que modificarlo, si se encuentra en un disco duro habrá que tenerse mucho cuidado en no borrar otros parámetros del sistema.

TRANS.PAS.- Este archivo funciona como enlace entre ADC.PAS y ADD.PAS, ya que muchos datos generados en ADC.PAS son utilizados por ADD.PAS. El primero escribe esos datos en el archivo TRANS.PAS, y el segundo los lee y completa su información con los datos de DATOS1.PAS.

Este archivo no debe ser modificado en ningún momento pues la información contenida en él se complementa, como ya se mencionó antes, con otro archivo de datos, y si ambos no se ajustan, no corre el programa ADD.PAS o sus resultados no son verídicos.

DATOS4.PAS.- En este archivo que es temporal se almacenan los resultados proporcionados por un subprograma de ADC.PAS, estos datos permiten a otro subprograma obtener la información necesaria para manipularla y llegar a resultados decisivos para la salida de resultados, es por esto que se recomienda no modificar el contenido de este archivo.

DATOS5.PAS.- Es otro archivo temporal en donde se guardan datos que son usados por ADC.PAS y ADD.PAS, por lo que podría también llamarse de enlace como TRANS.PAS aunque en menor grado.

No debe modificarse lo incluído en este archivo pues la información es necesaria para obtener las salidas adecuadas.

CORIN.PAS.- Es un archivo donde son guardadas las coordenadas de cada individuo; en el programa de análisis de correspondencias, este archivo no es usado, sin embargo, si se quisiera usar otro paquete para otros cálculos u otras gráficas, sería de mucha utilidad.

Las coordenadas de los individuos son conservadas si, y solo si, IWFI = 1 (IWFI es un parámetro que adquiere su valor en DATOS2.PAS).

CORVAR.PAS.- En este archivo son retenidas las coordenadas de cada variable, y como con CORIN.PAS, el archivo no es usado en ADC.PAS ni en ADD.PAS, pero podría ser necesario si se quisiera usar un paquete para más cálculos o para otro tipo de gráficas.

Las coordenadas de las variables son almacenadas solamente cuando IWFJ = 1 (IWFJ es un parámetro que toma su valor en DATOS2.PAS).

V EJEMPLOS DE APLICACION

V.1 caso de la marihuana por inhalación o ingerida

El siguiente ejemplo es con respecto al efecto de la marihuana por inhalación y por ingestión oral en cinco ratones de laboratorio durante siete días. Se pretende registrar el impacto que sufren los roedores con las dos modalidades de ingestión.

Este trabajo tiene una base teórica*, se toma el ejemplo para ser mostrado gráficamente utilizando el análisis factorial de correspondencias. Para este análisis se tienen:

Líneas o individuos: ratón 1(rat1), ratón 2 (rat2), ratón 3 (rat3), ratón 4 (rat4) y ratón 5 (rat5)

Columnas o variables:

1A : primer día : acción inmediata

1B : primer día : acción a los 10 minutos

1C : primer día : acción a la hora

2A : segundo día : acción inmediata

2B : segundo día : acción a los 10 minutos

2C : segundo día : acción a la hora

3A : tercer día : acción inmediata

.

7C : séptimo día : acción a la hora

Los datos serán ceros y unos, donde: el cero indica que la acción está inactiva y el uno que está activa.

A continuación se mostrarán los resultados proporcionados por el programa.

* Cordero González Guadalupe, García Sánchez Ma. Guadalupe y Piña Ballesteros Olivia "La marihuana" En *Muestra, Revista de la Escuela Nacional Preparatoria.* ep. 1, N° 3, enero-febrero-marzo, 1988. pp 41-46

EFECTOS DE LA MARIHUANA INHALADA EN RATONES

5 INDIVIDUOS DE LOS CUALES 0 SON SUPLEMENTARIOS

21 VARIABLES DE LOS CUALES 0 SON SUPLEMENTARIOS

2 FACTORES DE EXTRACCION

IMPRESION DE FACTORES SOBRE EL CONJUNTO DE INDIVIDUOS

IMPRESION DE FACTORES SOBRE EL CONJUNTO DE VARIABLES

1 GRAFICA(S) DESEADA(S)

NOM[J]	?	Di1A	Di1B	Di1C	di2A	Di2B	Di2C	Di3A	Di3B	Di3C	Di4A	Di4B	Di4C	Di5A	Di5B	Di5C	Di6A	Di6B	Di6C	Di7A	Di7B
PJ[J]	?	5	5	0	5	5	1	4	4	1	5	4	1	5	4	3	5	5	3	4	4
RAT1	?	1	1	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1
RAT2	?	1	1	0	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1
RAT3	?	1	1	0	1	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1
RAT4	?	1	1	0	1	1	0	1	1	0	1	0	0	1	1	0	1	1	0	1	1
RAT5	?	1	1	0	1	1	0	1	0	0	1	1	0	1	0	0	1	1	0	0	0

NOM[J]	?	Di7C
PJ[J]	?	3

EDICION DE VALORES PROPIOS

EL PRIMER VALOR PROPIO (PARASITO) ES ELIMINADO 1,0000000E+00

SUMA DE LOS VALORES PROPIOS ACTIVOS 2,86400565E-01

?	VALOR PROPIO	PORCENTAJE	PORCENTAJE ACUMULADO	HISTOGRAMA DE LOS PRIMEROS VALORES PROPIOS
1	1,04249295E-01	3,64E+01	3,64E+01	**
2	7,39551909E-02	2,58E+01	6,22E+01	**
3	6,19635765E-02	2,16E+01	8,39E+01	**
4	4,62254357E-02	1,61E+01	1,00E+02	**
5	0,00000000E+00	0,00E+00	1,00E+02	*
6	0,00000000E+00	0,00E+00	1,00E+02	*
7	0,00000000E+00	0,00E+00	1,00E+02	*
8	0,00000000E+00	0,00E+00	1,00E+02	*
9	0,00000000E+00	0,00E+00	1,00E+02	*
10	0,00000000E+00	0,00E+00	1,00E+02	*
11	0,00000000E+00	0,00E+00	1,00E+02	*
12	0,00000000E+00	0,00E+00	1,00E+02	*
13	0,00000000E+00	0,00E+00	1,00E+02	*
14	0,00000000E+00	0,00E+00	1,00E+02	*
15	0,00000000E+00	0,00E+00	1,00E+02	*
16	0,00000000E+00	0,00E+00	1,00E+02	*
17	0,00000000E+00	0,00E+00	1,00E+02	*
18	0,00000000E+00	0,00E+00	1,00E+02	*
19	0,00000000E+00	0,00E+00	1,00E+02	*
20	0,00000000E+00	0,00E+00	1,00E+02	*

26 PUNTOS EN EL PLANO DETERMINADO POR LOS EJES 1 Y 2

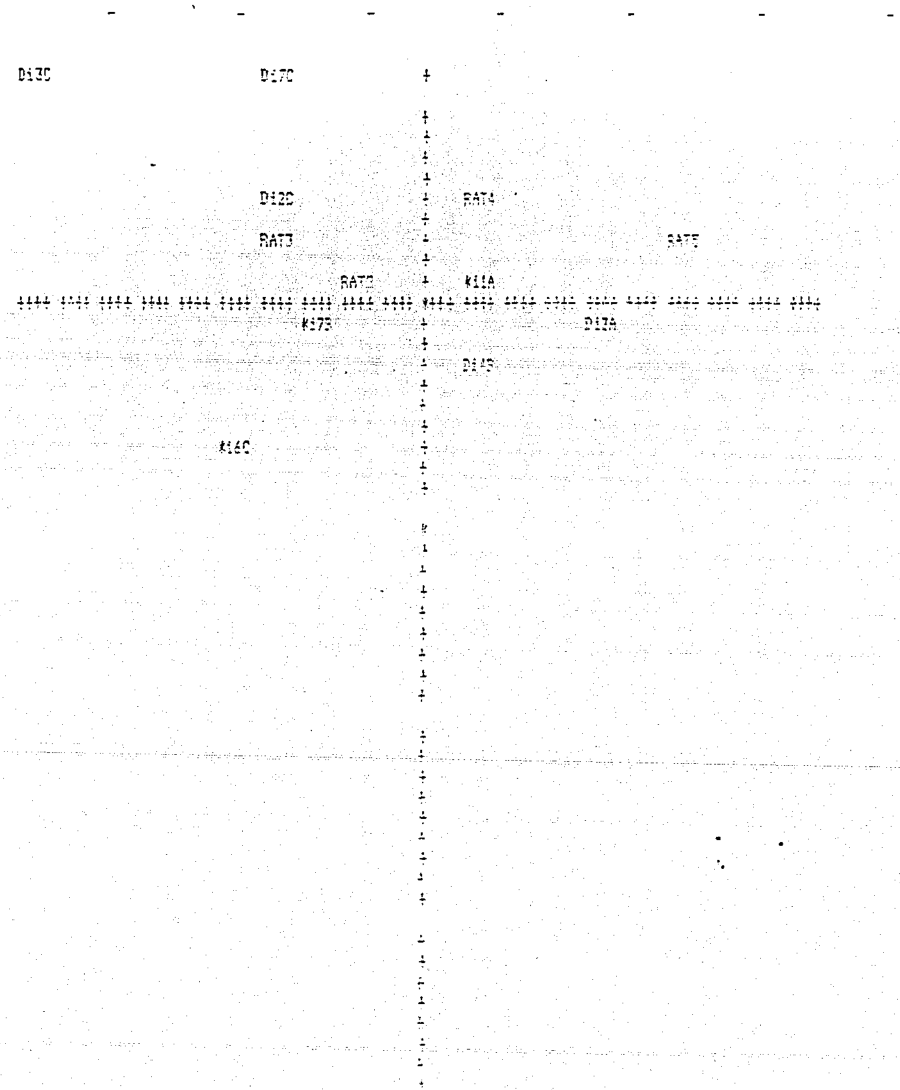
VALORES EXTREMOS SOBRE EL EJE 2 -1.82701060E+00 5.5704466305E-01
INTERVALO ENTRE DOS LINEAS 4.776882639E-02

```
I   D13C D17C
I---I
I   I
I   I
I   I
I   I
I   D12C RAT4
I   I
I   RAT3 RAT5
I   I
I   RAT2 D11A D11B d12A D12B D14A D15A D16B D16A
I---D11C ####
I   D17B D17A D13B D15B D13A
I   I
I   D14B
I   I
I   I
I   I
I   D16C D15C
I   I
I   I
I---I
I   RAT1
I   I
I   I
I   I
I   I
I   I
I   I
I   I
I   I
I---I
I   I
I   I
I   I
I   I
I   I
I   I
I   I
I   I
I   I
I---I
I   I
I   I
I   I
I   I
I   I
I   I
I   I
I   D14C
```

VALORES EXTREMOS SOBRE EL EJE 1 [illegible] 7.[illegible]
INTERVALO ENTRE DOS COLUMNAS [illegible]

EFECTOS DE LA MARIHUANA INHALADA EN RATONES

<u>EJES 1 Y 2</u>



Los datos anteriores fueron del consumo de marihuana por inhalación, los siguientes resultados son de cuando a los ratones se les administró la marihuana por vía oral. El estudio no fue completado en los siete días porque los ratones en el quinto día del experimento mostraron demasiada agresividad y dos de ellos murieron a causa de graves heridas.

EFECTOS DE LA MARIHUANA INGERIDA EN RATONES

5 INDIVIDUOS DE LOS CUALES 0 SON SUPLEMENTARIOS

15 VARIABLES DE LOS CUALES 0 SON SUPLEMENTARIOS

2 FACTORES DE EXTRACCION

IMPRESION DE FACTORES SOBRE EL CONJUNTO DE INDIVIDUOS

IMPRESION DE FACTORES SOBRE EL CONJUNTO DE VARIABLES

1 GRAFICA(S) DESEADA(S)

NOMEJI	?	D11A	D11B	D11C	D12A	D12B	D12C	D13A	D13B	D13C	D14A	D14B	D14C	D15A	D15B	D15C
PJCJI	?	4	3	2	0	5	5	0	1	4	0	5	4	2	5	5
RAT1	?	1	0	0	0	1	1	0	0	1	0	1	0	1	1	1
RAT2	?	1	1	1	0	1	1	0	1	1	0	1	1	0	1	1
RAT3	?	1	1	1	0	1	1	0	0	1	0	1	1	1	1	1
RAT4	?	0	0	0	0	1	1	0	0	0	0	1	1	1	1	1
RAT5	?	1	0	0	0	1	1	0	0	1	0	1	1	0	1	1

EDICION DE VALORES PROPIOS

EL PRIMER VALOR PROPIO (PARASITO) ES ELIMINADO 1,0000000E+00

SUMA DE LOS VALORES PROPIOS ACTIVOS 3,1344697E-01

?	? VALOR PROPIO	? PORCENTAJE	? PORCENTAJE ACUMULADO	? HISTOGRAMA DE LOS PRIMEROS VALORES PROPIOS
1	1,5029773E-01	4,79E+01	4,79E+01	***
2	1,0739871E-01	3,30E+01	8,10E+01	**
3	4,0464491E-02	1,29E+01	9,39E+01	********************
4	1,9242574E-02	6,14E+00	1,00E+02	********
5	0,0000000E+00	0,00E+00	1,00E+02	*
6	0,0000000E+00	0,00E+00	1,00E+02	*
7	0,0000000E+00	0,00E+00	1,00E+02	*
8	0,0000000E+00	0,00E+00	1,00E+02	*
9	0,0000000E+00	0,00E+00	1,00E+02	*
10	0,0000000E+00	0,00E+00	1,00E+02	*
11	0,0000000E+00	0,00E+00	1,00E+02	*
12	0,0000000E+00	0,00E+00	1,00E+02	*
13	0,0000000E+00	0,00E+00	1,00E+02	*
14	0,0000000E+00	0,00E+00	1,00E+02	*

EFECTOS DE LA MARIHUANA INGERIDA EN RATONES

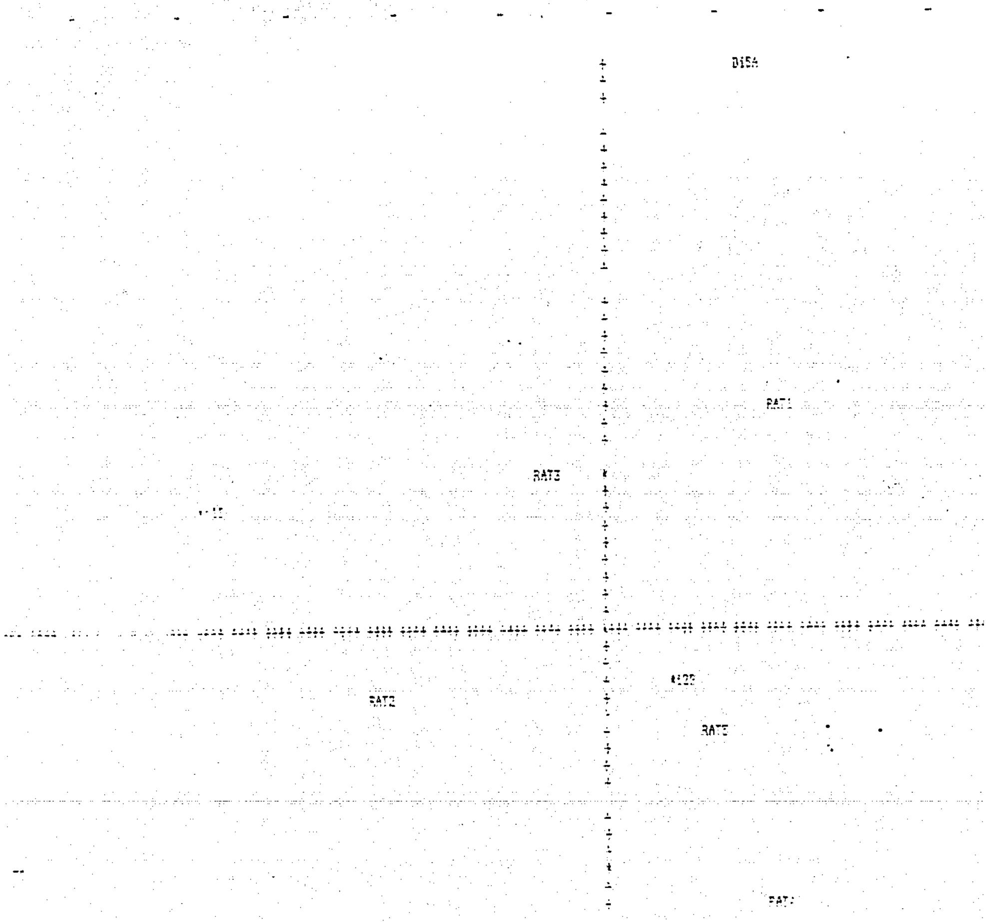
20 PUNTOS EN EL PLANO DETERMINADO POR LOS EJES 1 Y 2

VALORES EXTREMOS SOBRE EL EJE 2 -5.42193993E-01 1.082538716E+00
INTERVALO ENTRE DOS LINEAS 3.252714893E-02

```
I   D15A
I   I
I   I
I---I
I   I
I   I
I   I
I   I
I   I
I   I
I   I
I   I
I   I
I---I
I   I
I   I
I   I
I   I
I   I
I   I
I   RAT1
I   I
I   I
I---I
I   RAT3 D11A D15C
I   I
I   D11B D11C
I   I
I   I
I   I
I   I
I   I
I   I
I   I
I---D12A D13A D14A ++++
I   I
I   I
I   D12B D12C D13B D14B D15B
I   RAT2
I   I
I   RAT5
I   I
I   I
I   I
I---I
I   I
I   I
I   I
I   D13B D14C
I   I
I   RAT4
```

VALORES EXTREMOS SOBRE EL EJE 1 [illegible] [illegible]
INTERVALO ENTRE DOS [illegible] [illegible]

EFECTOS DE LA MARIHUANA INGERIDA EN RATONES
EJES 1 Y 2



Como resultado de este estudio a nivel de gráficas, se puede apreciar que en la primera de éstas, el quinto ratón está en la parte superior derecha del plano –en el primer cuadrante–, esto significa, considerando que el eje horizontal es el de mayor representación gráfica, que tuvo mucha actividad por la inhalación de la droga. El reporte del estudio revela que este ratón murió en el séptimo día. La gráfica también muestra que los ratones mantenian un mismo comportamiento, pues los puntos no están muy alejados unos de otros; mientras que en la segunda gráfica, esto es muy evidente, pues los animales se dispensan con respecto a sí mismos y con respecto a los días registrados. En este último caso, como se dijo, dos ratones murieron, aunque no por el consumo de la marihuana, sino por la agresión que se intensificó por el consumo de ésta.

V.2 Análisis idiomático de citas

Otro ejemplo es tomado del área bibliotecológica*, se trata de un análisis idiomático de las citas que cada artículo del *Annual Review of Energy* maneja, con el fin de observar la preponderancia de los idiomas en la publicación de artículos científicos, el análisis se hará para los artículos de todo un año. Las variables que se van a manejar son las siguientes:

Como líneas están el número de artículo: ar01, ar02, ar03, ...

Coma variables o columnas se tomarán: número de autores del artículo (# aut), el número total de referencias (refe) y el número de referencias según el idioma: Inglés (ingl), Ruso (ruso), Francés (fran), Español (espa), Alemán (alem), otros (otro) con esta categoría se pretende incluir a la mayoría de idiomas que pocas veces son citados, es decir, tienen poca literatura al respecto (energía).

Los resultados de la computadora al aplicar el análisis factorial de correspondencias son los siguientes:

* Meneses Tello Felipe. "Análisis idiomático de la literatura citada en los artículos científicos del *Annual Review of Energy*" En *Investigación Bibliotecológica.* Vol. 3, N° 6, enero-junio, 1989. pp. 9-21.

ANALISIS IDIOMATICO DE CITAS

30 INDIVIDUOS DE LOS CUALES 0 SON SUPLEMENTARIOS

8 VARIABLES DE LOS CUALES 0 SON SUPLEMENTARIOS

2 FACTORES DE EXTRACCION

IMPRESION DE FACTORES SOBRE EL CONJUNTO DE INDIVIDUOS

IMPRESION DE FACTORES SOBRE EL CONJUNTO DE VARIABLES

1 GRAFICA(S) DESEADA(S)

NOMEJE	?	#AUT	REFE	INGL	RUSO	FRAN	ESPA	ALEM	OTRO
PJEJE	?	29	499	671	0	0	24	3	1
AR01	?	1	12	12	0	0	0	0	0
AR02	?	1	[illegible]	27	0	0	0	3	0
AR03	?	1	74	74	0	0	0	0	0
AR04	?	2	42	42	0	0	0	0	0
AR05	?	3	94	94	0	0	0	0	0
AR06	?	2	39	39	0	0	0	0	0
AR07	?	2	21	21	0	0	0	0	0
AR08	?	2	44	44	0	0	0	0	0
AR09	?	2	29	29	0	0	0	0	0
AR10	?	1	27	3	0	0	24	0	0
AR11	?	1	32	32	0	0	0	0	0
AR12	?	1	23	23	0	0	0	0	0
AR13	?	1	35	35	0	0	0	0	0
AR14	?	2	24	24	0	0	0	0	0
AR15	?	1	54	54	0	0	0	0	0
AR16	?	1	8	7	0	0	0	0	1
AR17	?	1	21	21	0	0	0	0	0
AR18	?	1	3	5	0	0	0	0	0
AR19	?	2	47	57	0	0	0	0	0
AR20	?	1	25	25	0	0	0	0	0

EDICION DE VALORES PROPIOS

EL PRIMER VALOR PROPIO (PARASITO) ES ELIMINADO 1,00000000E+00

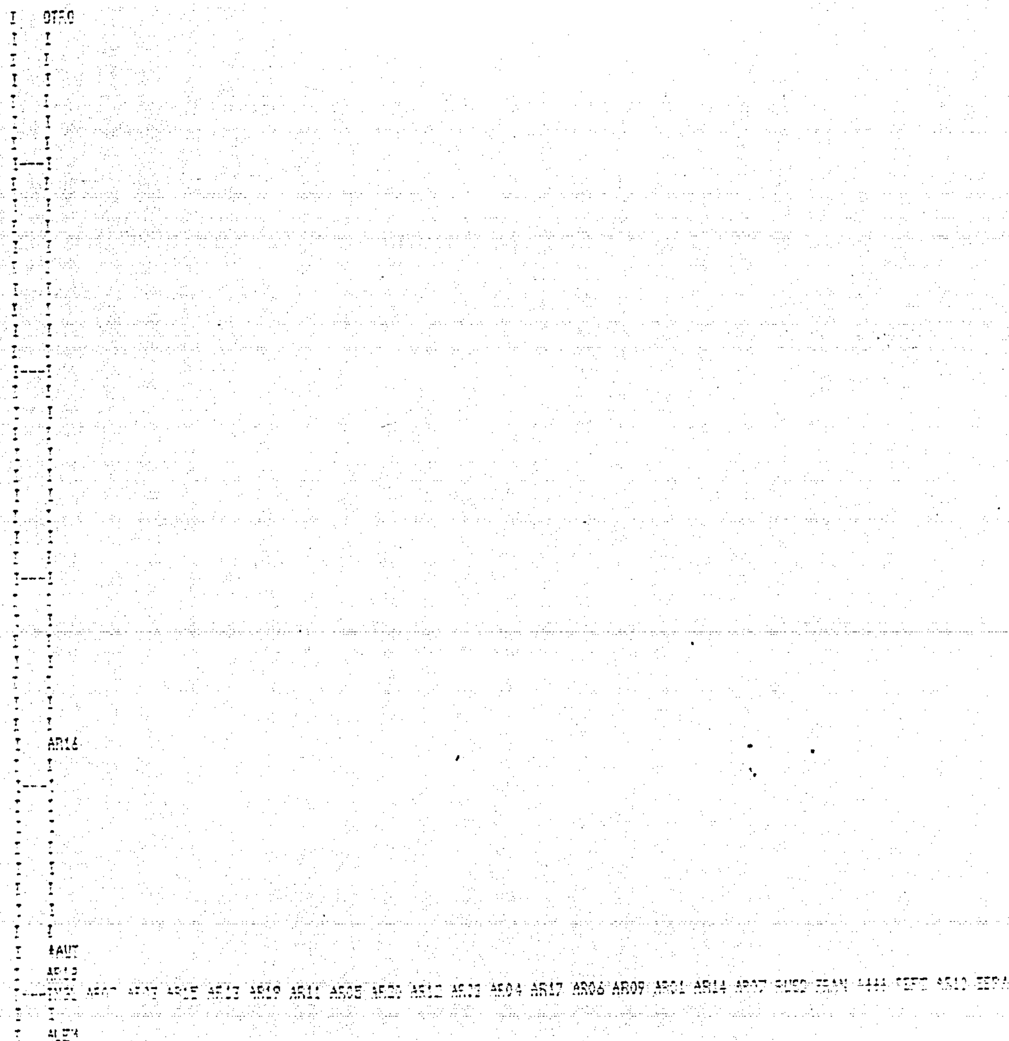
SUMA DE LOS VALORES PROPIOS OBTENIDOS [illegible]E-01

		VALOR PROPIO	PORCENTAJE	PORCENTAJE ACUMULADO	HISTOGRAMA DE LOS PRIMEROS VALORES PROPIOS
?	1	4,[illegible]E-01	7,96E+01	7,96E+01	**
?	2	[illegible]E-01	1,09E+01	9,04E+01	*********
?	3	4,[illegible]E-02	[illegible],43E+00	9,91E+01	********
?	4	5,[illegible]E-03	[illegible],22E-01	1,00E+02	*
?	5	0,00000000E+00	0,00E+00	1,00E+02	*
?	6	0,00000000E+00	0,00E+00	1,00E+02	*
?	7	0,00000000E+00	0,00E+00	1,00E+02	*

ANALISIS IDIOMATICO DE CITAS

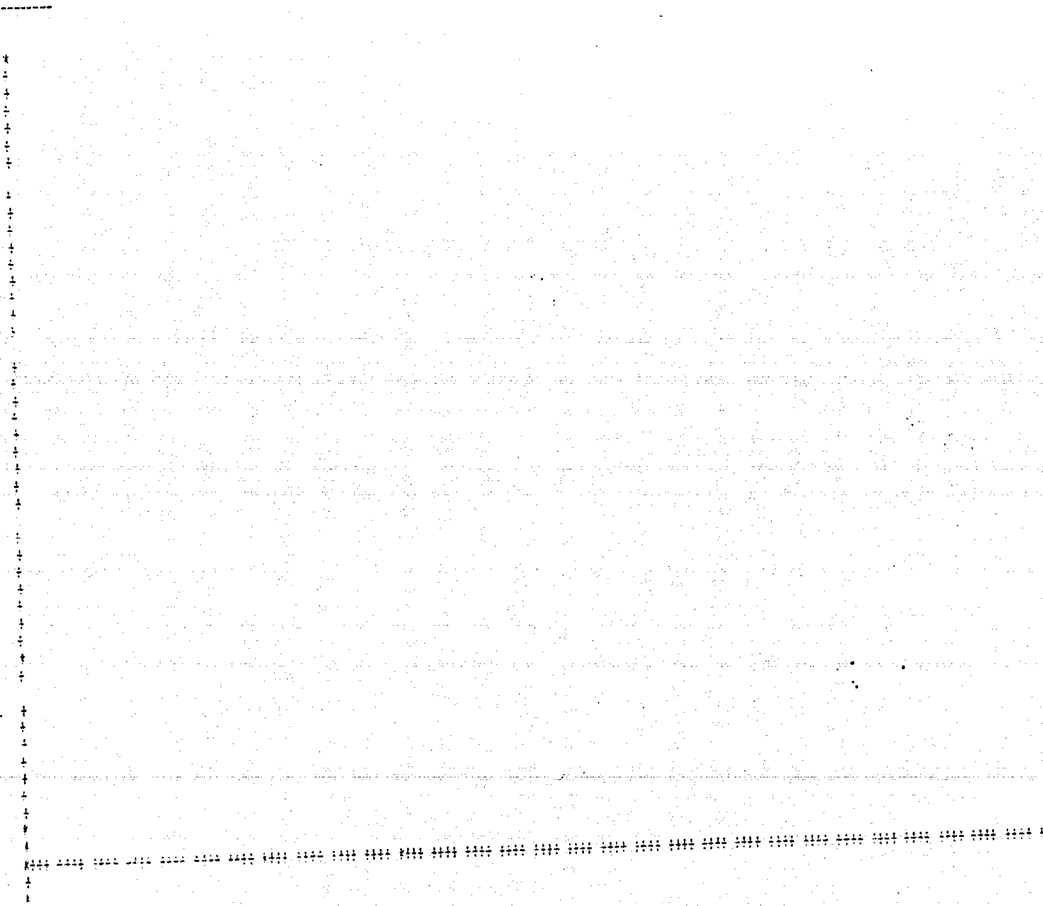
29 PUNTOS EN EL PLANO DETERMINADO POR LOS EJES 1 Y 2

VALORES EXTREMOS SOBRE EL EJE 2 -5.52535434E-01 9.0560182785+00
INTERVALO ENTRE DOS LINEAS 1.9240328535-01



VALORES EXTREMOS SOBRE EL EJE 1 [illegible] 4,9945344735+00
INTERVALO ENTRE DOS COLUMNAS [illegible]

ANALISIS IDIOMATICO DE CITAS
EJES 1 Y 2



Con este ejemplo se deduce fácilmente que el idioma que predomina es el inglés, es muy natural que así sea, considerando que se trata de artículos científicos y en esta área el poder idiomático es de la lengua inglesa. Sería interesante hacer una comparación con otras áreas del conocimiento en este mismo tópico y así determinar qué idioma es predominante para cada ciencia o materia.

Los datos fueron tomados del estudio citado para el año 1986, cubriendo muchos tipos de energía desde fuentes hasta utilización.

V.3 Educación de adultos en Texcoco

Para aplicar los elementos suplementarios se hizo una pequeña recaudación de datos referentes a la educación de adultos impartida en la Ciudad de Texcoco en el estado de México. Se pretende obtener un poco de información de las relaciones que guardan los individuos con respecto a algunas variables.

Las líneas serán los individuos entrevistados.

Las columnas tendrán las siguientes variables y categorías: Edad {15-24 años (eda1), 25-34 años (eda2), 35-44 años (eda3), de 45 años en adelante (eda4)}, Sexo {femenino (feme), masculino (masc)}, Etapa de educación {alfabetización (alfa), primaria (prim), secundaria (secu)}

La diferencia de este ejemplo con los anteriores es que aquí se van a emplear los elementos suplementarios, específicamente para este caso serán variables. Estas variables suplementarias se determinaron para sondear si los adultos que asisten al Instituto Nacional para la Educación de Adultos (INEA) en Texcoco, van voluntariamente o es exigencia de su trabajo o si necesitan aprender a leer y escribir (por ejemplo cartas de familiares) o tiene algún otro motivo. Para este fin se tienen las siguientes variables: van por voluntad propia (VOLU), van por razones de trabajo (TRAB), necesitan aprender algo específico (NECE) o tienen otros motivos (OTRO).

Los resultados de la computadora al aplicar el análisis factorial de correspondencias son los siguientes:

28 INDIVIDUOS DE LOS CUALES 0 SON SUPLEMENTARIOS

13 VARIABLES DE LOS CUALES 4 SON SUPLEMENTARIOS

2 FACTORES DE EXTRACCION

IMPRESION DE FACTORES SOBRE EL CONJUNTO DE INDIVIDUOS

IMPRESION DE FACTORES SOBRE EL CONJUNTO DE VARIABLES

VARIABLE(S) SUPLEMENTARIA(S) :volu trab nece otro

2 GRAFICA(S) DESEADA(S)

NOMEJJ	7	1524	2534	3544	45ad	feme	masc	alfa	pris	secu	volu	trab	nece	otro
PJEJJ	7	5	7	12	4	20	8	7	13	8	10	7	7	4
AD01	?	0	1	0	0	1	0	0	1	0	0	0	1	0
AD02	?	0	0	1	0	1	0	0	1	0	0	0	1	0
AD03	?	0	0	1	0	1	0	0	0	1	0	1	0	0
AD04	?	0	1	0	0	1	0	1	0	0	1	0	0	0
AD05	?	0	0	1	0	0	1	0	0	1	1	0	0	0
AD06	?	0	0	0	1	0	1	0	0	1	1	0	0	0
AD07	?	1	0	0	0	1	0	0	1	0	0	1	0	0
AD08	?	0	1	0	0	1	0	0	1	0	0	1	0	0
AD09	?	0	0	1	0	0	1	0	1	0	0	1	0	0
AD10	?	0	0	1	0	0	1	0	1	0	0	1	0	0
AD11	?	1	0	0	0	1	0	0	0	1	0	0	0	1
AD12	?	1	0	0	0	1	0	0	0	1	0	0	0	1
AD13	?	0	1	0	0	1	0	1	0	0	0	0	1	0
AD14	?	0	0	1	0	1	0	1	0	0	0	0	1	0
AD15	?	0	0	0	1	1	0	1	0	0	1	0	0	0
AD16	?	0	0	0	1	0	1	1	0	0	1	0	0	0
AD17	?	0	1	0	0	1	0	0	1	0	0	0	1	0
AD18	?	0	0	1	0	1	0	0	1	0	0	0	1	0
AD19	?	0	0	1	0	1	0	0	0	1	1	0	0	0
AD20	?	0	0	0	1	1	0	0	1	0	0	0	1	0
AD21	?	0	0	1	0	1	0	0	1	0	1	0	0	0
AD22	?	0	0	1	0	0	1	0	1	0	0	1	0	0
AD23	?	0	0	1	0	0	1	0	1	0	0	1	0	0
AD24	?	1	0	0	0	1	0	1	0	0	0	0	0	1
AD25	?	1	0	0	0	0	1	1	0	0	0	0	0	1
AD26	?	0	1	0	0	1	0	0	1	0	1	0	0	0
AD27	?	0	1	0	0	1	0	0	0	1	1	0	0	0
AD28	?	0	0	1	0	1	0	0	0	1	1	0	0	0

EDICION DE VALORES PROPIOS

EL PRIMER VALOR PROPIO (PASABITO) ES ELIMINADO = 1,0000000E+00

SUMA DE LOS VALORES PROPIOS ACTIVOS   2,0000000E+00

```
--------------------------------------------------------------------------------------------------------------------
?   ? VALOR PROPIO ? PORCENTAJE ? PORCENTAJE ? HISTOGRAMA DE LOS PRIMEROS VALORES PROPIOS
?   ?              ?            ? ACUMULADO ?
--------------------------------------------------------------------------------------------------------------------
? 1 ? 4,8001[illegible]E-01? 2,40E+01 ? 2,40E+01 ? **************************************************************************************
? 2 ? 4,5[illegible]E-01? 2,2[illegible]E+01 ? 4,65E+01 ? **********************************************************************************
? 3 ? 3,9[illegible]E-01? 1,96E+01 ? 6,64E+01 ? ************************************************************************
? 4 ? 2,849[illegible]E-01? 1,42E+01 ? 8,07E+01 ? ***************************************************
? 5 ? 2,05[illegible]E-01? 1,03E+01 ? 9,09E+01 ? *************************************
? 6 ? 1,817[illegible]E-01? 9,09E+00 ? 1,00E+02 ? *********************************
? 7 ? 0,0000000E+00? 0,00E+00 ? 1,00E+02 ? *
? 8 ? 0,0000000E+00? 0,00E+00 ? 1,00E+02 ? *
```

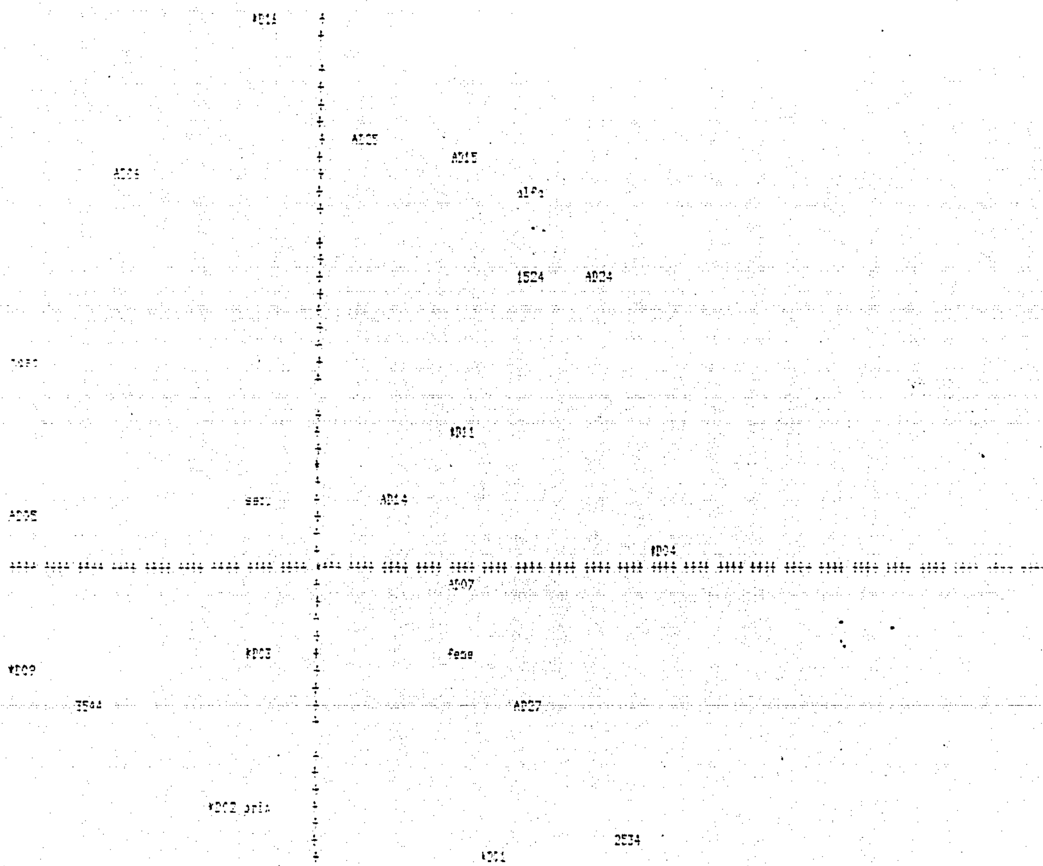
37 PUNTOS EN EL PLANO DETERMINADO POR LOS EJES 1 Y 2

VALORES EXTREMOS SOBRE EL EJE 2 -8.486036?2E-01 1.598932665E+00
INTERVALO ENTRE DOS LINEAS 4.899967785E-02

```
I   AD16 4Ead
I   I
I---I
I   I
I   I
I   I
I   I
I   AD23
I   AD15
I   AD01
I   alfa
I   I
I---I
I   I
I   I
I   1521 AD24
I   I
I   I
I   I
I   I
I   I
I   aasc
I   I
I---I
I   I
I   AD11 AD12
I   I
I   AD20
I   I
I   espc AD14
I   AD05
I   I
I   AD04 AD13
I---I
I   AD07
I   I
I   I
I   I
I   I
I   AD03 AD19 AD28 feme
I   AD09 AD10 AD22 AD23
I   I
I   IEA4 AD27
I   I
I---I
I   I
I   I
I   I
I   AD?? AD15 AD21 prim
I   I
I   IEZ4
I   AD?4 AD?? AD17 AD2?
```

VALORES EXTREMOS SOBRE EL EJE 1 -1.45?5151?E+00 1.11?94707?E+00
INTERVALO ENTRE DOS COLUMNAS [illegible]

Educaci'n de adultos en Texcoco
EJES 1 Y 2



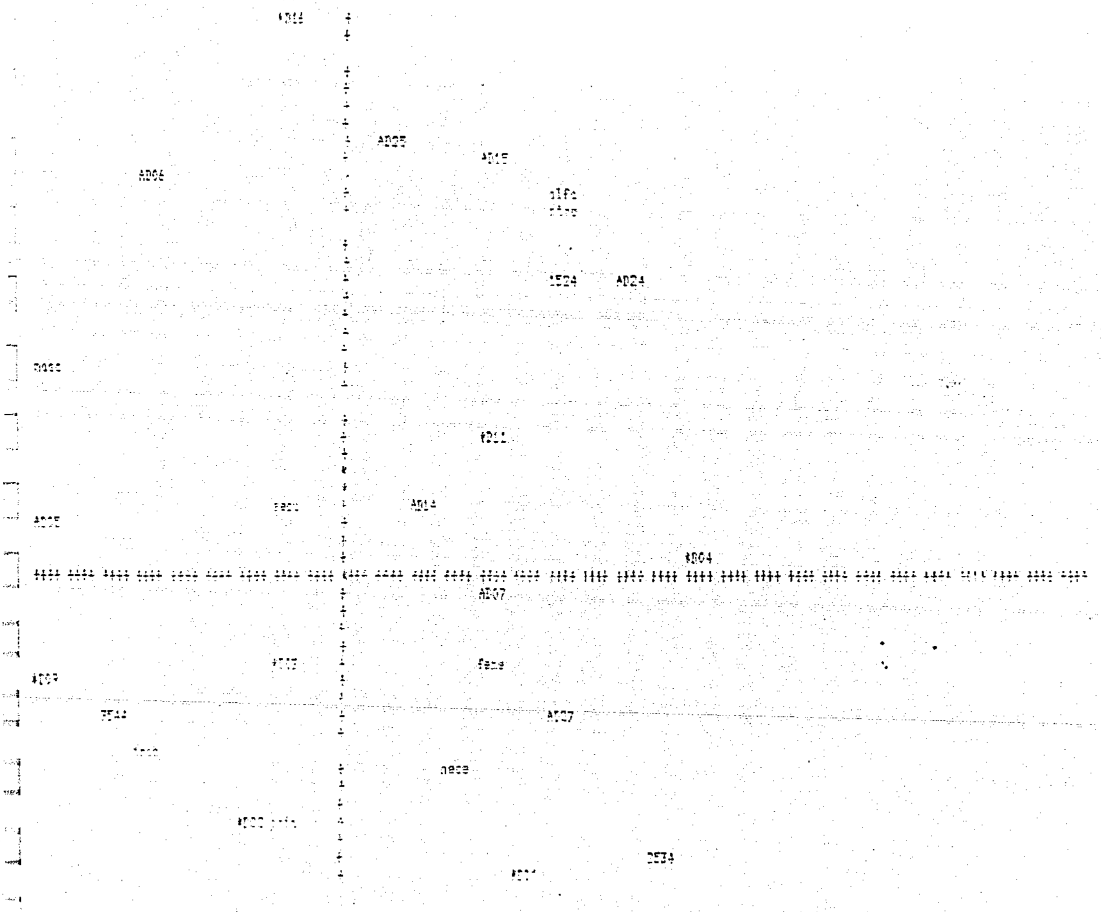
41 PUNTOS EN EL PLANO DETERMINADO POR LOS EJES 1 Y 2

VALORES EXTREMOS SOBRE EL EJE 2 -8.486036925E-01 1.598932665E+00
INTERVALO ENTRE DOS LINEAS 4.899967785E-02

```
I   AD16 45ad
I   I
I---I
I   I
I   I
I   I
I   I
I   AD25
I   AD15
I   AD06
I   alfa
I   otro
I---I
I   I
I   I
I   1524 AD24
I   I
I   I
I   I
I   I
I   nosi
I   I
I---I
I   I
I   AD11 AD12
I   I
I   AD??
I   volu
I   sec1 AD14
I   AD05
I   I
I   AD04 AD13
I---I
I   AD07
I   I
I   I
I   I
I   AD03 AD19 AD23 fema
I   AD09 AD10 AD22 A???
I   I
I   ES44 AD27
I   I
I---trab
I   nece
I   I
I   I
I   AD02 AD18 AD21 ????
I   I
I   2534
I   AD?? AD08 AD?? AD??
```

[illegible]
[illegible]

Educación de adultos en Texcoco
EJES 1 Y 2



En la primera gráfica solamente aparecen los elementos activos y en la segunda los activos y los suplementarios, en sí no hay diferencia, pero la información que adicionalmente proporcionan los suplementarios es valiosa. Por ejemplo, se deduce que hay una población mayor de mujeres que de hombres, la variable "feme" se encuentra más hacia la derecha; dada la concentración de adultos en la variable "prim", indica que hay más adultos en primaria que en los otros niveles; y finalmente por los datos adicionales, muchos van por voluntad a estudiar. Otro resultado relevante es la edad de los asistentes, la mayoría son de 35 a 44 años y además sus necesidades son de trabajo, esto es resultado de la cercanía de las variables correspondientes.

Sin embargo, los resultados se complementan con otras observaciones que se efectúan en el momento de las encuestas. Por ejemplo los resultados no muestran que los hombres no van solos, a los encuestados se les vio acompañados por otra persona que también hace uso de los cursos del INEA. Se aclara que la mayoría de las mujeres también asisten acompañadas.

En cuanto a los motivos de asistencia, los que manifestaron "otros", era para continuar los estudios a nivel medio superior en sistema abierto.

En esta encuesta no está planteada ninguna hipótesis y es difícil concluir algún aspecto; la intensión es solamente ejemplificar la utilización del análisis factorial de correspondencias.

CONCLUSIONES

Con esta obra se ha intentado mostrar la utilización del análisis de correspondencias, que por supuesto, puede ser empleado en cualquier área del conocimiento. Las ciencias sociales y humanísticas, utilizan con mayor frecuencia los diferentes métodos del análisis multivariado porque suelen ser los más aptos para aplicarse en todos los tipos de relaciones cuantificables, aunque con el análisis factorial de correspondencias las relaciones cualitativas no quedan fueran, es más, es especial para éstas.

Como todos los demás métodos de la estadística descriptiva, el análisis factorial de correspondencias no proporciona alguna parcialidad en la interpretación de los resultados, es completamente objetivo y es el usuario quien imprime la subjetividad a la interpretación. El análisis de correspondencias es una técnica descriptiva de datos que proporciona al usuario resultados con la mínima distorsión de graficación.

La modalidad que se maneja en esta obra, es la de los elementos suplementarios, los cuales, suministran información adicional de las tendencias de los datos activos (no suplementarios), facilitando la interpretación de los resultados. Con la utilización de este nuevo tipo de datos, la subjetividad del usuario al describir los resultados se reduce aún más y la contrastación de alguna hipótesis en alguna investigación es más fácil.

Los programas adicionados a este trabajo tienen portabilidad del 100%, la intensión de éstos es poder utilizarlo con el mínimo espacio posible y en cualquier computadora personal compatible con IBM, es opcional utilizar técnicas de multimedia, pero la implementación de esto queda a cargo del usuario.

BIBLIOGRAFIA

L'Analyse factorielle et ses applicationes. Centre National de la Recherche Scientifique, Paris, 1955.

ANDERSON, T.-W. *An Introduction to multivariate statistical analysis.* John Wiley & Sons, New York, 1958.

BENZECRI, J. P. *L'Analyse des données.* Dunod, Paris, 1973.

BURDEN, Richard L. y J. Douglas Faires. "Técnicas iterativas en el álgebra matricial" En *Análisis Numérico* tr. Simon Mochon C. México, D. F., Gpo. Editorial Iberoamérica, 1985. pp. 512-529.

BOUROCHE, Jean-Marie et Gilbert Saporta. "L'analyse factorielle des correspondances" En L'analyse des données. Saint-Germain, Presses Universitaires de France, 1980. pp. 83-108. (Que sais je?)

CASANOVA, Francisco y otros. "Contaminación de automóviles en la Ciudad de México" En *Ciencia y Desarrollo.* a. 12, num. 70, septiembre-octubre 1986. pp. 51-58.

CHATFIELD, Christopher and Alexander J. Collins. *Introduction to multivariate analysis.* Chapman and Hall, Londres, 1980.

CLAPAREDE, E. "Procedimientos auxiliares" En *La escuela y la Psicología,* tr. Ma. Luisa Navarro y Juan Comas. Losada, Buenos Aires, 1960.

DEMPSTER, A. P. *Elements of continuos multivariate analysis.* Addison-Wesley, Massachusetts, 1969.

DIDAY Edwin et Ludovic Lebart. "L'analyse des données" En *La Recherche,* Vol. 8, N° 74, janvier 1977. pp. 15-25

ELORZA, Harold. *Estadística para ciencias del comportamiento.* Harla, México, 1987.

GREENACRE, MICHAEL J. *Theory and applications of correspondence analysis.* Academic Press, Londres, 1984.

HARRIS, Richard J. *A primer of multivariate statistics.* Academic Press, New York, 1975.

EGGHE, L. and R. Rousseau. *Informetrics 87/88.* Select Proceedings of the Firts International Conference on Bibliometrics and Theoretical Aspects of Information Retrieval Diepenbeek, Bélgica, 25-28 de agosto, 1987.

JACKSON, BARBARA B. *Multivariate data analysis : an introduction.* Richard D. Irwin, Illinois, 1983.

KALBFLEISCH, J. D. *Probability and statistical inference.* Springer Verlang, Berlin, 1979.

LEBART, Ludovic. *Multivariate descriptive statistical analysis : correspondance analysis and related techniques forlarge matrices.* John Wiley & Sons, New York, 1984.

MARDIA, K. V. et al. *Multivariate analysis.* Academic Press, Londres, 1979.

Mc GREGOR ANCIOLA, Federico. *Analyse des structures économico-industrielles et des contraintes associées au développement des pays dans la définition des programmes énergétiques : une proposition méthodologique.* L'Agence Internationale de l'Energie Atomique, Viena, 1984.

MORA y ARAUJO, Manuel y otros. *El análisis de datos en la investigación social.* Nueva Visión, Buenos Aires, 1973. (Tratados Fundamentales)

OSTLE, Bernard. *Estadística aplicada.* LIMUSA, México, 1981.

PANIAGUA, Mónica G. *El análisis de correspondencias : un método de análisis multivariado descriptivo.* México, 1986.

SZEKELY, Béla. "Los test y el análisis factorial" En *Los Tests.* ed. 4ª, vol. I, Buenos Aires, Kapelusz. pp. 153-177.

TIJSSEN, R. J. W. *Scientometric Data-Analysis Methods : an introduction to correspondence analysis for complete and incomplete tables.* Centre for Science and Technology Studies : University of Leiden, 1989.

——— et al. "Quasi-correspondence analysis on scientometric transaction matrices" En *Scientometrics.* vol. II, nos. 5-6, 1987. pp. 351-366.

WANG, Peter C. C. *Graphical representation of multivariate data.* Academic Press, New York, 1978.

WIRTH, Niklaus and Kathleen Jensen. *Pascal : user manual and report.* Springer-Verlag, Berlin, 1978.

A mis padres:

Silvia y Antonio (q. p. d.)

A mis hijos:

Fanny, Angel y Jazmín

A mi esposa:

Margarita

A mis hermanos

Isidro, Marco A. y José Luis

A la Sra.:

Fanny Zamudio de C.

A mi amigo y hermano:

Gabriel Curiel Z.

A la familia:

Curiel Zamudio

A mis amigos:

Arturo Bazán Z.

Alejandro Benavides Z.

Eréndira Valdez C.

Raúl Villaseñor G.

Mi gratitud al Dr. Hugo Rincón M.,
por su estímulo y amistad.

Agradezco a todos mis profesores y compañeros su amistad y enseñanzas, en especial al Dr. Alejandro J. Díaz Barriga C. por su comprensión y paciencia en la dirección de esta tesis. Asimismo a la Profa. Lydia López Amador por sus consejos y apoyo.

A mis compañeros de la U.P.N.