

# **Modelo de análisis de correspondencias múltiples**

Javier Enrique Parra Olivares

Centro de Estadística e Investigación de Operaciones.

Facultad de Ciencias Económicas y Sociales.

La Universidad del Zulia, Maracaibo, Venezuela.

Teléfono: +58 61 596568-596528. E-mail: jparra@luz.ve

## **Resumen**

En este artículo se presenta, de manera muy abreviada, una de las técnicas del análisis multivariable de datos: el Modelo de Análisis de Correspondencias Múltiples, la cual tiene la virtud de que ha sido desarrollada para estudiar el tipo de variable más frecuente en la investigación social de campo: las variables cualitativas o de nivel de medición nominal. Así mismo, se muestra uno de los procedimientos computacionales que permite aplicar este modelo a grandes matrices de datos: el procedimiento CORMU del Système Portable pour L'analyse des Données (SPAD.N).

**Palabras clave:** Análisis multivariable, correspondencias múltiples, variables cualitativas, programa SPAD.N

## **Multiple Correspondences Analysis Model**

### **Abstract**

The present article shows, in a very brief way, one of the multivariate data analysis techniques: Multiple Correspondences Analysis Model, that has been developed in order to study the most frequent type of variable in the empirical social research: the qualitative one.

At the same time, it shows one of the computational routines that allows to apply this technique to very large data matrix: Routine CORMU of the Système Portable pour L'analyse des Données (SPAD.N).

**Key words:** Multivariate Analysis, Multiples Correspondances, Qualitative variables, SPAD.N Software.

## **1. Introducción**

Las Ciencias Sociales, en la última década, han vuelto su atención a las investigaciones de tipo exploratorio, tal vez impulsadas por la desconfianza en los niveles de explicación de los modelos teóricos tradicionales.

Esta tendencia ha coincidido, a su vez, con el re-descubrimiento de modelos estadístico-matemáticos que posibilitan el manejo simultáneo de muchas variables para explorar los rasgos más relevantes de cualquier fenómeno o situación. El renacer de estas técnicas se ha visto facilitado por el arrollador desarrollo de la informática, lo que ha permitido elaborar programas de computación que hacen posibles cálculos, transformaciones y representaciones gráficas en muy corto tiempo, solventando las principales desventajas prácticas que presentaba el uso de estas herramientas de análisis de datos.

## **2. Finalidad y conceptos del modelo de análisis de correspondencias múltiples**

Una de las técnicas estadísticas multivariable de más reciente desarrollo es el ANALISIS DE CORRESPONDENCIAS MULTIPLES. La materia prima empírica para este tipo de técnica la constituye una matriz de datos de orden  $n \times p$ , con  $n$  individuos medidos a través de  $p$  variables, de las cuales la mayoría alcanza sólo un nivel de medición nominal. Esto último en el sentido expuesto por S. S. Stevens en 1951, cuando definió a la medición como “la asignación de numerales de acuerdo con reglas” y, en el caso de la escala nominal, “la regla es: no asignar el mismo numeral a diferentes clases o diferentes numerales a la misma clase” (Stevens et al, 1976; 65).

Esta técnica, enmarcada dentro de los métodos estadísticos factoriales, está diseñada para estudiar las relaciones entre cualquier número de modalidades de distintas variables categóricas. En esto presenta dos claras diferencias con respecto a otro método factorial, el Análisis de Componentes Principales, que estudia: a) las relaciones entre variables y no entre valores de éstas y b) que se nutre de variables de tipo cuantitativo.

El Análisis de Correspondencias Múltiples utiliza unos cálculos de ajuste que recurren esencialmente al álgebra lineal y produce unas representaciones gráficas donde los objetos a describir se transforman en puntos sobre un eje o en un plano. Su principal aplicación, en la actualidad, se halla en el campo del análisis exploratorio de datos, en el cual se aprovecha "...la capacidad descriptiva del método, sobre todo en su aspecto disociador de categorías de diferentes variables cuando éstas son suficientes en número para hacer difícil un estudio global de ellas". (Sánchez Carrión, 1984; 99).

Desde el punto de vista matemático, las consideraciones que se realizan al interior de esta técnica son de carácter geométrico, encontrándose dentro de los procedimientos descriptivos de la estadística; por lo tanto, el tipo de resultados que se alcanzan tiene un carácter meramente descriptivo del colectivo estudiado, por lo cual es particularmente "...idóneo para aplicarlo a situaciones donde sean pocas o ninguna las hipótesis previas de trabajo y se requiera un análisis exploratorio de la situación a tratar (a través de una muestra del colectivo en estudio) con el fin de establecer los puntos de partida de análisis posteriores." (Sánchez Carrión, 1984; 75).

El procedimiento en sí consiste en construir una matriz simétrica formada por  $Q^2$  bloques denominada **Tabla de Burt**, la cual contiene, en los bloques de la diagonal principal, las tablas de frecuencia (producto del cruce de una variable consigo misma) de cada una de las  $Q$  variables categóricas de interés y fuera de la diagonal principal las tablas de contingencia resultantes del cruce de cada uno de los pares de variables consideradas relevantes (Ver Ejemplo en Tabla I).

El paso siguiente se ejecuta a través de la diagonalización de la matriz asociada, lo cual conduce a los autovalores (eigenvalues), valores característicos o valores propios asociados con cada uno de los factores o ejes extraídos.

La matriz a diagonalizar es de la forma que se especifica a continuación:

$$V = \frac{1}{Q} D^{-1} B$$

donde:

Q: número de variables consideradas relevantes.

D: **matriz diagonal cuyos elementos diagonales** son los de la Tabla de Burt, los efectivos o frecuencia absoluta de cada modalidad. El resto de los elementos son nulos (o ceros).

B: Tabla de Burt.

Tabla 1  
Tabla de Burt  
(Ejemplo con 3 Variables Categóricas Activas)

i	VARIABLE 1						VARIABLE 2						VARIABLE 3								
	ACOM	ANOV	AMUS	AVAR	APEL	AHUM	RCOM	RNOV	RMUS	RVAR	RPEL	RINF	RHUM	R???	SMCO	SMNO	SMPL	SMMU	SMHU	SM??	
ACOM	38	0	0	0	0	0															
ANOV	0	32	0	0	0	0															
AMUS	0	0	6	0	0	0															
AVAR	0	0	0	38	0	0															
APEL	0	0	0	0	10	0															
AHUM	0	0	0	0	0	4															
RCOM	0	3	0	3	2	0	8	0	0	0	0	0	0	0							
RNOV	10	10	1	11	2	0	0	34	0	0	0	0	0	0							
RMUS	2	0	0	1	0	0	0	0	3	0	0	0	0	0							
RVAR	9	8	2	13	1	3	0	0	0	36	0	0	0	0							
RPEL	4	1	0	0	0	0	0	0	0	0	5	0	0	0							
RINF	0	2	2	4	2	0	0	0	0	0	0	10	0	0							
RHUM	4	4	0	0	0	0	0	0	0	0	0	0	8	0							
R???	9	4	1	6	3	1	0	0	0	0	0	0	0	24							
SMCO	25	24	4	20	5	1	8	20	3	19	3	4	3	19	79	0	0	0	0	0	0
SMNO	2	2	0	2	1	0	0	3	0	3	0	0	0	1	0	7	0	0	0	0	0
SMNU	2	1	0	3	1	0	0	2	0	3	0	1	1	0	0	0	7	0	0	0	0
SMPL	2	0	2	3	2	0	0	2	0	2	0	4	1	0	0	0	0	9	0	0	0
SMIN	2	4	0	5	0	3	0	3	0	8	1	0	1	1	0	0	0	0	14	0	0
SMHU	3	1	0	3	0	0	0	1	0	0	1	1	2	2	0	0	0	0	0	7	0
SM??	2	0	0	2	1	0	0	3	0	1	0	0	0	1	0	0	0	0	0	0	5

Los autovalores o valores característicos ( $\mu\alpha$ ) asociados con la matriz anterior on de la forma:

$$\mu\alpha = \frac{1}{nQ} \phi\alpha B \phi\alpha \quad (2)$$

onde:

$\phi\alpha$ : autovector que permite establecer las proyecciones de las modalidades sobre cada eje del análisis.

n: número de individuos analizados.

Cada autovalor  $\mu$  mide la suma de las proyecciones de las modalidades sobre el eje  $\alpha$  (Morineau et al, 1985; 283-284). Por lo tanto, el autovalor es una medida de la variabilidad captada por el factor construido por el análisis, ya que se pueden interpretar como nuevas variables "artificiales" o rasgos "subyacentes" al conjunto original de datos, posibilitando así una mejor comprensión, teórica y/o empírica según el caso, de las interrelaciones existentes entre las distintas modalidades de las variables consideradas.

Habiéndose determinado cada factor, las proyecciones ortogonales de las diferentes modalidades sobre el mismo permiten establecer cuáles son las modalidades que mayor aporte o contribución realizaron en la formación del factor. De allí se define como **Contribución Absoluta** a la proporción de varianza "explicada" por un factor debida a una modalidad.

Así mismo, otro coeficiente de gran importancia para el análisis es el denominado **Contribución Relativa** que es la correlación entre modalidad y factor en este contexto. En esa medida expresa la contribución de un factor en la "explicación" de la dispersión de una modalidad o, lo que es lo mismo, es la proporción de la distancia de la modalidad al centro de gravedad que es explicada por un factor dado (Ver Ejemplo en Tabla II). De esa manera, "...mientras que las contribuciones absolutas permiten saber qué variables son las responsables de la construcción de un factor, las contribuciones relativas muestran cuáles son las características exclusivas de este factor". (Morineau et al, 1985; 318).

Un paso final, en cuanto al uso de la técnica, consiste en la posibilidad de asignarle una "definición" o concepto al factor o eje, de acuerdo con las modalidades que más aportaron a su formación y con las que mejor se "dejaron representar" por el mismo. Este paso, que no siempre es posible por razones teóricas y/o empíricas, requiere un mínimo de claridad teórico-metodológica en cuanto a la naturaleza del fenómeno analizado, los objetivos del estudio y la habilidad analítica del investigador.

**Tabla II**  
 (Ejemplo de Tabla de Coordenadas, Contribuciones Absolutas y  
 Contribuciones Relativas para 5 Variables y 5 Ejes o Planos de Análisis).  
 Coordenadas, Contribuciones Absolutas y Relativas de las Modalidades Activas en los Ejes 1 a 5

MODALIDADES	COORDENADAS					CONTRIBUCIONES ABSOLUTAS					CONTRIBUCIONES RELATIVAS					
	PREL	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
23 ACEPT																
ACOM - ACEPTA COMIQUITAS	5.94	27	13	04	-67	-78	1.1	3	0	8.5	11.6	0.1	0.1	0.0	1.9	2.6
ANOV - ACEPTA NOVELAS	5.00	87	-28	-27	36	34	9.3	1.1	1.1	2.0	1.8	2.5	0.3	0.2	0.4	0.4
AMUS - ACEPTA MUSICALES	94	-1.61	-65	33	1.38	-23	6.0	1.1	3	5.7	2.1	1.3	0.2	0.1	0.9	0.0
AVAR - ACEPTA VARIOS	5.94	5	17	17	17	21	1.1	1.8	0	7	8.1	1.2	0.5	0.0	0.1	0.2
APEL - ACEPTA PELICULAS	1.56	-66	-1.31	24	-0.2	0.7	1.7	7.4	0	1.7	0.4	0.5	0.1	0.0	-0.1	-0.4
AHUM - ACEPTA HUMORISTICOS	63	-41	2.01	88	12	1.41	3	6.9	1.4	0	1.9	0.1	1.3	0.2	0.0	0.6
CONTRIBUCION ACUMUL																
- 22.3 18.6 3.1 16.8 20.6																
24 RECHZ																
RCOM - RECHAZA COMIQUITAS	1.25	12	-1.12	1.00	-3.5	6.2	0	4.3	3.6	5	1.6	0.0	0.8	0.7	0.1	0.3
RNOV - RECHAZA NOVELAS	5.31	11	-2.9	-8.0	-3.8	6.2	2	1.2	5.5	2.5	6.6	0.0	0.3	1.3	0.5	1.4
RMUS - RECHAZA MUSICALES	4.7	-1.99	0.5	5.6	-2.60	-9.4	1.4	0	4	10.0	1.3	0.3	0.0	0.1	1.6	0.2
RVAR - RECHAZA VARIOS	5.63	-0.6	9.9	3.5	3.5	2.4	1	15.2	2.0	2.2	1.1	0.0	3.8	0.5	0.5	0.2
RPEL - RECHAZA PELICULAS	7.8	1.02	2.6	-1.03	-1.58	-1.36	2.0	1	2.4	6.2	4.6	0.4	0.0	0.4	1.0	0.8
RINF - RECHAZA INFORMATIVOS	1.56	-1.56	-1.07	-8.7	1.24	-3.0	9.5	4.9	3.4	7.6	4	2.1	1.0	0.6	1.3	0.1
RHUM - RECHAZA HUMORISTICOS	1.25	8.4	-0.6	-1.02	7.1	-1.35	2.2	0	3.8	2.0	7.3	0.5	0.0	0.7	0.3	1.2
R*** - RECHAZA SIN INFORMAC	3.75	1.9	-2.9	8.4	0.3	-8.4	3	9	7.6	0	2.8	0.1	0.2	1.6	0.0	0.5
CONTRIBUCION ACUMUL																
= 15.7 26.8 28.7 31.0 25.7																
25 HOR8M																
8MCO - COMIQUITAS 8 AM	12.34	1.9	-2.0	2.7	-1.9	-0.8	1.1	1.3	2.7	1.4	3	0.6	0.6	1.2	-0.6	0.1
8MNO - NOVELAS 8 AM	1.09	1.10	-3.9	4.8	4.8	1.06	3.3	5	7	8	3.9	0.7	0.1	0.1	0.1	0.6
8MMU - MUSICALES 8 AM	1.09	-2.8	5.9	-1.10	6.1	-7.9	2	1.0	3.8	1.3	2.2	0.0	0.2	0.7	0.2	0.4
8MPL - PELICULAS 8 AM	1.41	-2.01	-7.4	-7.3	7.2	-3.2	14.1	2.1	2.2	2.3	8	3.1	0.4	0.4	0.4	0.1
8MIN - INFORMATIVOS 8 AM	2.19	0.8	1.72	-1.9	5.0	9.1	0	17.8	2	1.7	5.8	0.0	3.6	0.0	0.3	1.0
8MHU - HUMORISTICOS 8 AM	1.09	1.7	-5.5	-6.8	2.3	-1.51	1	9	1.5	2	8.0	0.0	0.2	0.3	0.0	1.3
8M?? - NO CONTESTO 8 AM	7.8	-9.6	1.3	-6.6	-1.53	1.27	1.8	0	1.0	5.8	4.0	0.4	0.0	0.2	1.0	0.7
CONTRIBUCION ACUMUL																
= 20.6 23.7 12.1 13.5 24.9																
26 HOR4P																
4PCO - COMIQUITAS 4 PM	7.34	-3.3	-0.5	-7.9	1.2	-0.2	2.0	1	13.2	4	0	0.6	0.0	3.6	0.1	0.0
4PNO - NOVELAS 4 PM	3.44	-5.5	-9.5	9.0	3.4	4.4	2.6	8.6	8.0	1.3	2.1	0.6	1.9	1.7	0.2	0.4
4PMU - MUSICALES 4 PM	3.91	-0.6	4.0	7.8	0.8	-8.3	0	1.8	6.9	-1	8.6	0.0	0.4	1.5	0.0	1.7
4PPE - PELICULAS 4 PM	2.66	-1.13	-2.0	-2.8	-8.5	-5.5	8.4	3	6	6.1	2.6	2.0	0.1	0.1	1.1	0.5
4PIN - INFORMATIVOS 4 PM	1.09	-5.6	7.2	3.4	9.1	-6.2	8	1.6	2	2.9	1.3	0.2	0.3	0.0	0.5	0.2
4PHU - HUMORISTICOS 4 PM	1.56	-2.9	1.16	0.9	-7.4	7.0	3	5.8	0	2.7	2.4	0.1	1.1	0.0	0.5	0.4
CONTRIBUCION ACUMUL																
= 14.2 18.2 29.0 13.4 17.0																
27 HOR8P																
8PCO - COMIQUITAS 8 PM	3.13	-2.0	-4.2	2.4	-1.9	-4.3	3	1.5	5	13.9	1.8	0.1	0.3	0.1	2.6	0.3
8PNO - NOVELAS 8 PM	6.09	-4.9	-3.7	3.8	6.5	2.1	3.6	2.3	2.6	8.1	8	1.0	0.6	0.6	1.8	0.2
8PMU - MUSICALES 8 PM	1.25	1.59	-2.0	-1.92	-5.9	-1.1	7.8	1	13.4	1.4	1	1.7	0.0	2.5	-0.2	0.0
8PPE - PELICULAS 8 PM	3.91	-1.14	-0.5	-3.3	-0.2	-0.3	12.5	0	1.2	0	0	3.1	0.0	0.3	0.0	0.0
8PIN - INFORMATIVOS 8 PM	9.1	9.9	1.13	-1.26	7.8	8.6	2.3	3.3	4.4	1.8	2.2	0.5	0.6	0.8	0.3	0.4
8PHU - HUMORISTICOS 8 PM	4.06	-2.5	-5.8	2.0	0.1	-6.6	6	3.7	5	0	5.6	0.2	0.8	0.1	0.0	1.1
8M?? - NO CONTESTO 8 PM	6.1	3.7	1.01	1.59	-3.3	-7.9	2	1.7	4.6	2	1.2	0.0	0.3	0.8	0.0	0.2

La anterior fue una exposición muy sucinta de los principales conceptos del Análisis de Correspondencias Múltiples desde el punto de vista estadístico y metodológico. Es preciso resaltar que, aunque los principios subyacentes a los aspectos técnicos de este método tienen antecedentes lejanos - Hirschfeld en 1935, Fischer en 1940 y Benzecri en 1964, entre otros - (Morineau et al, 1985; 276), el desarrollo de la técnica alcanza grandes posibilidades con el surgimiento y generalización de los computadores y con los trabajos de Benzecri.

### **3. Técnica computacional: procedimiento "CORMU" del sistema portable de análisis de datos (SPAD.N)**

El paquete computacional denominado *Système Portable Pour L'analyse des Données* (SPAD) surgió en principio como una biblioteca de sub-rutinas implementadas en el lenguaje FORTRAN IV, que formaban parte de la obra "Techniques de la description statistique, méthodes et logiciels pour l'analyse des grands tableaux" (L. Lebart, A. Morineau y N. Tabard, 1977). Luego, ha venido siendo enriquecido con aportes de muchos investigadores del área de análisis de datos hasta conformar las versiones actuales. Está concebido como un conjunto de programas para el tratamiento estadístico integral de grandes tablas o matrices de datos, abarcando etapas de organización, cálculo y presentación de resultados analíticos y gráficos. (Noguera, 1987).

Desde la versión 1.00 (SPAD.N) para microcomputadores (1987), se incluye el procedimiento CORMU, que permite la aplicación del modelo de Análisis de Correspondencias Múltiples a matrices de datos de grandes dimensiones.

La cadena de procedimientos que permite desarrollar un Análisis de Correspondencias Múltiples es la siguiente: **ARDON - ARDIC - SELEC - CORMU - GRAPH**. En la descripción que sigue se empleó como referencia al Manual del Usuario del SPAD.N. (Centre International de Statistique et d'Informatique Appliquées, 1988).

El procedimiento **ARDON** permite cargar la matriz de datos brutos en el sistema, estableciendo el número de variables **p** a utilizar y el número de individuos a analizar (con su respectivo identificador o código).

El procedimiento **ARDIC** carga el diccionario de variables en el sistema. En esta fase se incluye cada variable en el orden en que aparecen en la matriz de datos, se fija el nombre de cada una y se declara su nivel de medición: nominal (o categórica) o continua (ordinal o intervalar). En el contexto del SPAD se reserva el término

“continuo” para las variables numéricas, independientemente de que sean discretas o continuas; en todo caso este tipo de variables siempre cumple una función ilustrativa o suplementaria dentro de la técnica de Análisis de Correspondencias Múltiples. Por otro lado, para las variables nominales se deben declarar los identificadores o etiquetas para cada modalidad, empleando un máximo de 4 caracteres.

El procedimiento **SELEC** organiza la información anterior (ARDON y ARDIC), estableciendo la correspondencia entre el diccionario y la matriz de datos. En esta fase se especifica a cuál clase de variables pertenece cada una de las mismas de acuerdo con su importancia en el análisis estadístico propuesto y en función de la siguiente tipología:

**Variables activas:** son las que participan directamente en las diferentes fases del cálculo, en especial en la formación de los factores.

**Variables ilustrativas:** no son tomadas en cuenta en ciertas fases del cálculo, pero se usan para establecer criterios de comparación en los sistemas de coordenadas factoriales.

**Variables abandonadas:** no participan de ninguna forma en los cálculos. Al no ser declaradas en el Procedimiento SELEC, el sistema las abandona aunque permanezcan en la matriz de datos y en el diccionario de variables.

El procedimiento **CORMU** es aquel que permite el tratamiento de las modalidades de las variables declaradas como activas en el procedimiento SELEC, con la finalidad de producir un análisis de correspondencias múltiples. Esta etapa incluye la construcción e impresión de la Tabla de Burt, así como la constitución de los factores para el análisis.

Finalmente, el procedimiento **GRAPH** edita los gráficos de coordenadas factoriales para el análisis de correspondencias múltiples e imprime las tablas de Contribuciones Absolutas y Relativas para cada factor.

#### **4. Análisis de los resultados**

Como producto de la ejecución del programa según los términos expresados en el aparte anterior, se obtienen, básicamente, los resultados siguientes:

- a) La Tabla de Burt (Ver Tabla I).
- b) Un histograma con los primeros 51 autovalores y los respectivos porcentajes de inercia o variabilidad que recogen del total (Ver Ejemplo en la Figura 1).



Figura 1  
(Ejemplo de Histograma de Autovalores)  
Histograma de los 29 primeros valores propios

NUMERO	VALOR PROPIO	PORCENTA.
1	.4037	6.96
2	.3623	6.25
3	.3441	5.93
4	.3168	5.46
5	.3137	5.41
6	.2886	4.98
7	.2817	4.86
8	.2748	4.74
9	.2550	4.40
10	.2492	4.30
11	.2263	3.90
12	.2195	3.79
13	.2105	3.63
14	.2014	3.47
15	.1902	3.28
16	.1819	3.14
17	.1726	2.98
18	.1599	2.76
19	.1497	2.58
20	.1419	2.45
21	.1270	2.19
22	.1195	2.06
23	.1153	1.99
24	.1080	1.86
25	.0963	1.66
26	.0821	1.42
27	.0760	1.31
28	.0681	1.17
29	.0642	1.11

- c) Un cuadro con las Coordenadas, Contribuciones Absolutas y Contribuciones Relativas (en los primeros cinco ejes) de las modalidades de las variables activas (Ver Tabla II).
- d) Una tabla con las Coordenadas y Valores-Test para las modalidades de las variables tanto activas como ilustrativas.
- e) Los gráficos representativos de los planos formados por las combinaciones de los primeros cinco ejes del análisis (Ver Ejemplo en la Figura 2).

A partir de dichos resultados, es posible analizar las interrelaciones entre las distintas modalidades de las variables categóricas consideradas, con la finalidad de explorar las dimensiones empíricas del fenómeno a estudio. Para tal fin, la estrategia del análisis consiste en considerar cada eje por separado, su inercia (variancia), las modalidades con mayor aporte a su formación y las que mejor estén captadas por cada eje, intentando luego, en función de lo anterior, conceptualizar a los ejes. En una última fase se consideran los planos formados por los distintos ejes conceptualizados, para así enriquecer el análisis.

Una primera decisión se debe tomar en esta etapa en relación a cuál es el número de ejes que se va a analizar. En este sentido el hecho es que no hay una forma objetiva de decidir cuántos ejes se deben retener, lo cual es una desventaja no sólo de esta técnica sino de todos los métodos factoriales. Sin embargo, algunos autores proponen ciertas reglas empíricas para auxiliarse en la toma de esta decisión.

Una de estas reglas empíricas es sugerida por Chatfield y Collins (1978) y consiste en "...observar el patrón de los autovalores y notar si existe un quiebre natural". Esta regla aparece más claramente formulada por Abascal y Fernández (1989, pág. 46): "Si representamos el histograma de los valores propios con los números de los ejes en ordenadas y los porcentajes de inercia explicada en abscisas, se pueden eliminar los ejes cuyo orden es posterior al 'codo' que se produce en la curva...". (Ver Ejemplo en la Figura 3).

Es preciso resaltar que para el análisis de los ejes se pueden considerar como importantes todas las modalidades que posean una contribución absoluta mayor al valor que determina la siguiente regla empírica:

$$CTA \geq \frac{2}{j} \times 100 \quad (3)$$

donde:

j: número total de modalidades activas para el estudio.

Así mismo, se pueden considerar sólo aquellas modalidades con frecuencias relativas mayores al 5%, también como regla empírica.

Figura 2  
 Ejemplo de Plano para 2 Ejes Factoriales

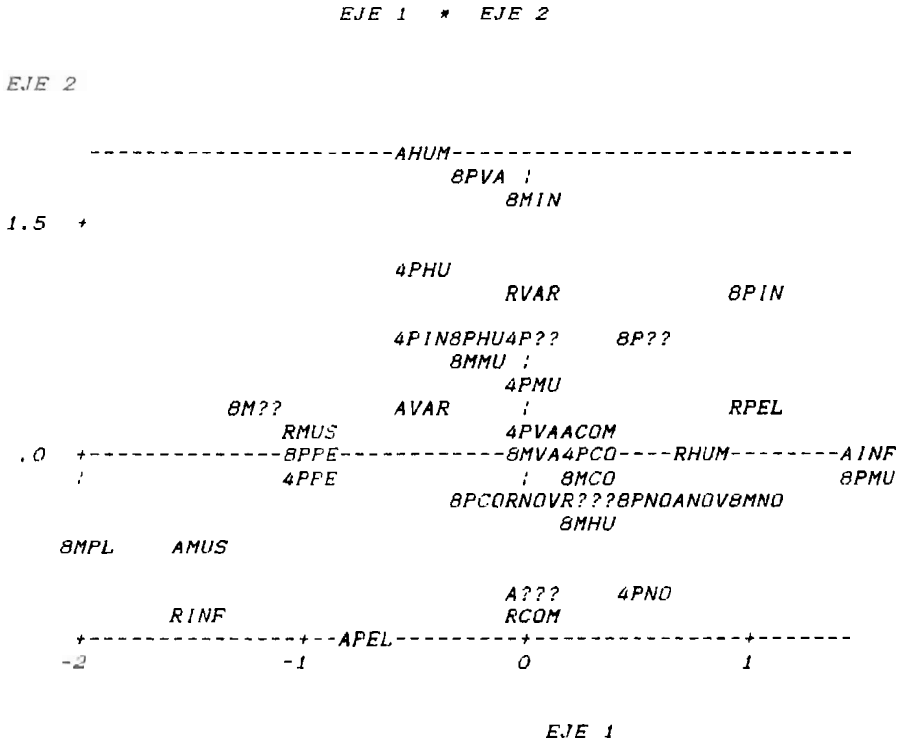


Figura 3  
Ejemplo de la Regla Empírica de Chatfield y Collins)  
Histograma de los 29 primeros valores propios

NUMERO	VALOR PROPIO	PORCENTA.
1	.4037	6.96
2	.3623	6.25
3	.3441	5.93
4	.3168	5.46
5	.3137	5.41
6	.2886	4.98
7	.2817	4.86
8	.2748	4.74
9	.2550	4.40
10	.2492	4.30
11	.2263	3.90
12	.2195	3.79
13	.2105	3.63
14	.2014	3.47
15	.1902	3.28
16	.1819	3.14
17	.1726	2.98
18	.1599	2.76
19	.1497	2.58
20	.1419	2.45
21	.1270	2.19
22	.1195	2.06
23	.1153	1.99
24	.1080	1.86
25	.0963	1.66
26	.0821	1.42
27	.0750	1.31
28	.0681	1.17
29	.0642	1.11

- En el autovalor 11 se produce el primer "quebre" importante.  
- Por tanto, se pueden elegir los primeros 10 para el análisis.

Para intentar una definición de cualquier eje es válido considerar a alguna de las modalidades de las variables declaradas como ilustrativas, de manera que permitan darle mayor significado al eje a analizar. Ahora bien, no todas las modalidades ilustrativas tienen alta importancia en determinados ejes, por lo que se requiere un criterio para tomar decisiones en este sentido.

Uno de los criterios a considerar tiene que ver con los denominados "valores-test". Una modalidad se puede considerar importante en un eje dado, si ella es obtenida de un conjunto de individuos con proyecciones altas o con contribuciones importantes en la formación de tal eje y que, por lo tanto, se alejan del origen de coordenadas.

Los valores-test son indicadores que permiten decidir qué tanto se aleja una modalidad en relación al origen de coordenadas. Requieren algunos supuestos muy rigurosos como normalidad en los datos y aleatoriedad en la selección de la muestra para asegurar la independencia entre sí de las proyecciones de los individuos. Ante la dificultad en la verificación de la normalidad, se puede apelar al teorema del límite central para asumir como normales (con esperanza matemática igual a cero y varianza el inverso de la frecuencia de la modalidad) a las distribuciones de aquellas modalidades que son escogidas por más de 30 individuos.

El valor-test se define por la siguiente ecuación:

$$\text{Valor - test} = \sqrt{nj} \phi \alpha_j \quad (4)$$

donde:

$n_j$ : número de individuos que seleccionó la modalidad  $j$ .

y se asume que si el valor absoluto del valor-test de una determinada modalidad es suficientemente alto, a un nivel de significación dado, es posible rechazar la hipótesis nula de que el promedio (esperanza matemática) de las proyecciones de los individuos que escogieron esa modalidad es igual a cero. De esta forma, si el valor-test, en términos absolutos, es mayor que el valor crítico pre-fijado, se considera como importante a la modalidad asociada (Lebart et al, 1977; 137).

## Referencias bibliográficas

### Libros

- ABASCAL, Elena y FERNÁNDEZ, Ildefonso. (1989). *Métodos Multivariantes para la Investigación Comercial*. Barcelona, Editorial Ariel.
- CENTRE INTERNATIONAL DE STATISTIQUE ET D'INFORMATIQUE APPLIQUÉES. (1988). *Manual del Usuario del SPAD.N*. Saint-Mandé.

- CHATFIELD C. y COLLINS, A. J. (1978). *Introduction to Multivariate Analysis*. Londres, Chapman and Hall.
- LEBART, L., MORINEAU, A. y TABARD, N. (1977). *Techniques de la description statistique, méthodes et logiciels pour l'analyse des grands tableaux*. Paris, Editorial Dunod.
- MORINEAU, A., LEBART, L. y FENELON, J. P. (1985). *Tratamiento Estadístico de Datos*. Barcelona-México, Marcombo-Boixareu Editores.
- NOGUERA, Carlos. (1987). *Manual con ejemplos del Spad*. Universidad Central de Venezuela, Mimeo.
- SANCHEZ CARRION, Juan Javier y Otros. (1984). *Introducción a las Técnicas de Análisis Multivariadas aplicadas a las Ciencias Sociales*. Madrid, Centro de Investigaciones Sociológicas.
- STEVENS, S. S. y otros (1976). *Escalas de Medición en las Ciencias Sociales*. Buenos Aires, Ediciones Nueva Visión.

**Tesis, Seminarios, Informes Técnicos.**

- CENTRE INTERNATIONAL DE STATISTIQUE ET D'INFORMATIQUE APPLIQUEE. (1988). *Manual del Usuario del SPAD*. N. Saint-Mandé.
- NOGUERA, Carlos. (1987). *Manual con ejemplos del Spad*. Universidad Central de Venezuela, Mimeo.