CONSIDERACIONES METODOLOGICAS Y EXPERIMENTALES DEL ANALISIS FACTORIAL EN GEOGRAFIA

F. FERNANDEZ GUTIERREZ

SUMMARY: The present article is a systematic commentary on the properties and use of the mathematical-statistical technique known as factorial analysis, having already shown the efficiency of quantitative techniques in the social science field, and determined their real possibilities. The commentary is based on concrete examples from the author's own geographical experience of the original data, the geographical nature of the technique, and improvement systems.

RESUME:
Concrétisée et démontrée l'efficacité des techniques quantitatives dans le domaine des sciences sociales, une fois déterminées et calibrées leur authentiques possibilités après les critiques et considérations qui suivirent l'etape expensive, le présent article est un comentaire méthodologique des propriétés et du mode d'emploi de la technique matheématico-statistique connue comme Analyse Factoriell. Tel commentaire s'effectue a partir de cas concrets extraits de l'expérience de l'auteur dans le domaine de la géographie. Il traite des aspects du dessin et contenu des matrices, du caractère spacial de la technique et de systèmes de perfectionnement et de dépuration.

I. INTRODUCCION.

Hoy día es generalmente aceptado el hecho de que a toda exigencia por alcanzar un nivel cada vez más científico en cualquier disciplina, le acompaña, de forma casi paralela, una necesidad de aumentar la cuantificación y matematización en sus técnicas metodológicas y en el tratamiento de la información.

Dentro del amplio espectro de las ciencias humanísticas, la Geografía como otras muchas han retrasado considerablemente el dar el paso cualitativo hacia una profundización y consolidación como ciencia adulta. Y ello, queramos o no, requiere cuantificar, ampliar el campo de trabajo con variables y unidades de observación; operar con modelos lógicos matemáticos estadísticos muy experimentados que ofrecen magníficas posibilidades a la Geografía tanto inductiva como deductiva.

La ofensiva renovadora de la Geografía y la introducción de los métodos cuantitativos para la manipulación de los datos pretende como uno de sus objetivos más inmediatos eliminar la parte negativa del inherente subjetivismo y personalismo, consciente o inconsciente, que contienen muchos estudios de nuestra materia.

Aceptando como notas supremas y esenciales de la nueva corriente la síntesis y la formulación de leyes, se plantea la necesidad de considerar simultáneamente toda una gama de múltiples factores, de los que hay que determinar la relación que se dan entre ellos y evaluar sus dependencias. Para ello, y todo lo anterior, el A.F. se ha mostrado como un método estadístico muy propio y experimentado en otras ciencias sociales afines de la Geografía como la Psicología, Ecología, Sociología, Economía, etc.

Esta técnica matemáticoestadística, se descubre a finales del Siglo XIX, pero su aplicación se pospone obligatoriamente hasta la aprición y desarrollo de la Cibernética en el presente siglo. La complejidad de su cálculo numérico lo hizo hasta entonces imposible. Fundamentalmente se basa en la idea de la correlación y en la determinación del factor principal (1). Algunos autores comparativamente, le dan actualmente un valor e importancia similar o más valiosos que el alcanzado por la propia cartografía dentro del campo específicamente geográfico. De ser así, y lo creemos firmemente, pasará a ser una técnica central nueva de la Geografía actual y futura.

1.- Desde hace bastantes años ha venido siendo empleado también por los geografos, sobre todo en Estados Unidos en trabajos que exigían el manejo de un considerable número de variables y unidades de análisis al mismo tiempo, relacionandolas todas simultaneamente en conjunto.

En contraste con los estudios en que se consideraban un menor números de casos y variables, conjugandolas entre si aisladamente, par a par o en grupos.

El Análisis Factorial (A.F.) es una técnica de análisis estadístico que detecta las dimensiones latentes, implícitas entre variables, no directamente accesibles a una observación necesariamente superficial (2).

El propósito del A.F. es determinar, a partir de las interrelaciones de un gran número de variables, el menor número posible de factores cuya asociación con las variables originales les haga significativo de todas las interrelaciones derivadas.

2.- Un factor equivalente a una supervariable que es en cierto modo el resultado de la unión de varias variables vinculadas e influenciadas entre si. Su interacción define un fenómeno con entidad propia. Si todas las correlaciones entre un gran número de variables pueden explicarse en términos de asociaciones de cada una de ellas con unos pocos factores como variables fundamentales, entonces estos factores pueden considerarse como variables, resumen y síntesis del conjunto inicial. Los factores principales coinciden normalmente con un número no determinado de las variables originales. Las correlaciones de las variables originales con los factores se conocen como cargas del factor (3).

Finalmente posibilita al geógrafo encontrar mediante los "pesos o scores" las diferencias espaciales de los principales factores y descubrir los tipos de combinaciones espaciales, según las cuales el sistema analizado se estructura y se organiza.

La propia cuantificación y el A.F. en concreto poseen unas exigencias y unas limitaciones propias. Desde los inicios de la investigación se requiere un tratamiento y una ordenación de la información. Hay que delimitar las unidades de observación (en lo sucesivo U.E.A. o U.O.), procurando que sean lo más homogéneas posible en todas sus características. En función de estas unidades se toman las variables o atributos espaciales en un cuestionario previamente codificado y estructurado o bien en hojas de datos de ordenador para facilitar su trasvase posterior a fichas.

El número de U.E.A. y el de variables está en relación de la capacidad del programa de A.F. que se vaya a usar en el ordenador (4). Independientemente del tipo de programa que utilicemos, nunca el número de variables puede rebasar al de U.O.; como máximo se podrá llegar a una cantidad igual. Es decir, "N" unidades de observación por "M" atributos. Lo más correcto es, que a ser posible no se llegue a apurar a tope esta norma ni tampoco la capacidad de tratamiento del programa, en cuanto a estas dos exigencias.

II. ETAPAS DEL ANALISIS FACTORIAL.

Una vez especificado el significado y las exigencias mínimas derivadas de la aplicación de ésta técnica, vamos a detenernos en analizar las propiedades y los aspectos que caracterizan el desarrollo de su proceso de ejecución bajo un punto de vista formal sin introducirnos en las propiedades de su base matemática.

1. *Matriz de información espacial.*

El punto de partida del A.F. es la configuración de la Matriz de Información Espacial (M.I.E.). La elaboración de la M.I.E. implica concretar con la máxima precisión las unidades de observación, medir los valores que en ellas alcanzan las distintas variables utilizadas y la selección correcta y previa de dichos atributos espaciales. La matriz está pues integrada por los atributos o variables y la U.E.A. (ver gráfico n.º 1). Las U.O. se delimitan realizando nosotros mismos la división o adoptando cualquier otra distribución efectuada sin nuestra intervención con fines administrativos, censales o estadísticos: Secciones, distritos, barrios, regiones, municipios, provincias, departamentos, etc.

Lo más correcto y recomendable es levantar las U.E.A. con arreglo a nuestros intereses, teniendo en cuenta la exigencia de una máxima homogeneidad en sus características comunes: tamaño superficial, tamaño cronológico, número de elementos: habitantes, edificios, etc.

Pero si recurrimos a las fuentes indirectas (estadísticas, publicaciones, archivos, etc.) probablemente nos encontramos obligados a tomar las divisiones ya efectuadas por los que recabaron la información, y que puede coincidir o no con nuestros fines. Este sería el caso de los Censos Agrarios de Población, Viviendas y Edificios. La excesiva heterogeneidad de la U.O. pueden originar dificultades en la comparación de las características. A fin de evitar los inconvenientes que a efectos estadísticos se pueden derivar, lo mejor es recurrir a expresar los valores de los atributos en cifras relativas (tantos por ciento) o realizar una normalización. En última instancia, aunque no es recomendable, pueden ir en cifras absolutas. En este sentido el A.F. tiene una cierta elasticidad, ya que el paso inicial que le sigue a la construcción de la M.I.E. es efectuar una estandarización de los valores incluidos en ella.

CUADRO 1

VARIABLES

Unidades de Observación	1 Solteros	2 Casados	3 Viudos	4 Separados	5 0 - 4 años	6 0 - 19 años	7 20 - 59 años	8 60 y +	9........ nacidos locales	70 - - -
Sección 1	54'8	42'5	2'7	0'5	10'5	41'6	46'8	1'1	68'5	- - - -
" 2	50'3	40'5	8'5	0'7	11'6	42'3	43'8	2'3	60'7	- - - -
" 3	49'7	44'9	5'3	0'1	12'3	45'3	39'2	3'2	63'3	- - - -
" 4	50'7	41'2	7'8	0'3	14'1	45'7	38'1	2'1	54'7	- - - -
" 5	52'5	43'5	3'5	0'7	9'8	39'3	45'6	5'3	49'8	- - - -
" 6	54'1	40'7	4'8	0'4	6'5	28'7	63'5	1'3	57'3	- - - -
" 7	45'1	48'7	5'8	0'4	6'5	28'7	63'5	1'3	57'3	- - - -
" 8	46'7	41'7	11'-	0'6	12'6	43'8	43'1	0'5	65'3	- - - -
" 9	49'3	40'3	9'9	0'5	11'2	42'7	45'5	0'7	75'4	- - - -
" .	.	.	.	.	.	.	.	.	.	.
" .	.	.	.	.	.	.	.	.	.	.
" 120	.	.	.	.	.	.	.	.	.	.

Cuando comentabamos las limitaciones del A.F. aludimos a la capacidad de la M. I. E. Si en algún caso se reabasase el tope de un mayor número de variables respecto al de UU.OO. o al programa usado, se tendría que recurrir a efectuar varios AA.FF. de forma encadenada. En el cuadro nº 1, ofrecemos un aspecto parcial de una M.I.E. de un estudio demográfico urbano que contaba con 120 observaciones y 70 variables. Las conclusiones metodológicas han sido extraidas de sendos AA.FF. de morfología y estructura urbana así como de estudios de habitat.

Para la selección de las variables y con el objeto de facilitar esta labor con la claridad que exige, se debe comenzar por definir, en hipótesis, la existencia de una serie de tramas o interconexiones de variables que influyen en el objeto de estudiar con un carácter más o menos independiente.

Cada trama tiene un número de atributos igual o inferior al número de U.E.A., como exige este tipo de cálculo matricial. En un estudio urbano, por ejemplo, podríamos partir de la existencia de varias tramas: Física, morfológica, demográfica, estructural, etc. Cada una de ellas poseerá una cantidad de variables específicas, que por su contenido se vincularán con la trama y la definirán. La medida de variables puede representar todos los aspectos de un país, región, municipio, ciudad, etc, altitud, clima, vegetación, población, actividades productivas, aspectos sociales, opiniones políticas o bien puede considerar un aspecto social, demográfico o económico.

Los elementos de la M.I.E., (Unidades de observación y atributos espaciales), es la información que se perfora en fichas del ordenador e integran los datos de base. Todas las secciones o demarcaciones llevan un número igual de tarjetas y la cantidad de éstas depende del número de variables de la matriz y de la cantidad de dígitos que se use.

2. *Matriz de Correlaciones.*

El primer paso del proceso es calcular el índice de correlaciones "Pearson product moment" entre cada par de variables.

Tras la tipificación de los datos a fin de que los valores de las variables puedan ser comparativos, se obtiene la matriz de correlaciones, que es una matriz cuadrada R:(m x m). El índice de correlación es una medida que precisa la dependencia lineal entre dos variables $(x_1\ x_2)$ consideradas (5). Al ser cuadrada, en ella se manifiestan las dependencias lineales —si existen— entre todas las variables entre sí, par a par. Mediante los índices de correlaciones pueden extraerse ya consideraciones interesantes a partir de la estructura de agrupación que se considere. El nivel de correlación puede ser descrito en términos numéricos, usando un índice de correlación "product moment". El índice "r" se encuentra siempre entre + 1 y — 1. Cuando el grado de correlación es elevado se registra un valor aproximado a 1 ó a — 1 si la correlación es negativa. Una correlación baja se sitúa en torno a 0.

R: +1; Indica una dependencia lineal creciente y positiva entre dos variables dadas x_1 y x_2. Es decir, que cuando aumenta por ejemplo X_1 lo hace también x_2 en la misma proporción y viceversa. El mismo juego se realiza cuando disminuye.

R: 0; Muestra que no hay correlación, que no existe ningún tipo de nexo ni de relación. Toda aproximación a cero gradua la intensidad de la independencia de X_1 y x_2.

R: —1; Significa una dependencia máxima lineal decreciente negativa entre x_1 y x_2. O sea, que cuando aumenta X_1 y X_2 disminuye en un porcentaje similar al aumento de la primera variable y viceversa.

siendo "m" las variables o atributos. En la figura geométrica resultante la diagonal contiene la correlación de la variable consigo misma siendo por lo tanto su valor 1. En la M. C. los coeficientes de correlación se repiten y son idénticos a uno y otro lado de la diagonal. A continuación añadimos un aspecto parcial de una M. C. de un estudio para una mejor comprensión (Cuadro n.º 2).

En la muestra anterior están subrayados los coeficientes de correlación significativos. Estos se obtienen mediante un test de significación en los que el coeficiente de correlación aumenta conforme el número de U.O. disminuye. Si el tivamente con la población de 60 años y más (—0,43).

Al incrementarse los niños y jóvenes de 0 a 19 años, aumenta en un mismo conjunto humano considerado los niños de 0 a 4, los nacidos en la localidad y los estudiantes de Primaria y desciende por el contrario el número de viejos de 60 y más años. Elevando el coeficiente de correlación de cada uno, da el porcentaje de dependencia entre ellos (r = 67%) para los niños de 0—4, etc. En el caso negativo al aumentar los individuos de 0 a 19 años en 100, descienden los viejos en un 18,5%. Cosa lógica ya que cuanto más jóven es una población el

CUADRO 2

MATRIZ DE CORRELACIONES (aspecto parcial)

›parados	0 - 4	0 - 19	20 - 59	60 y +	Nac. Locales	Nac. Provinc.	And. Orien.	And. Occid.	Resto España	Extranjero	Alfabetos	Analfabetos	Estd. Prima.	Est. Bach. Elemt.	Bach. Super.	Univer-sitar.	Adul. 1ª in-compl.	Adul. 1ª compl.	A. Bach. Elemt.	A. Bach. Sup. y univr.
.70	0.06	0.21	0.12	-0.14	0.18	0.04	0.21	0.10	0.20	0.05	0.30	-0.10	0.15	0.35	0.27	0.13	0.00	0.12	0.15	0.1
.90	0.05	0.02	0.38	-0.08	0.01	0.39	0.12	0.09	0.16	0.08	0.45	-0.24	-0.04	0.23	0.17	0.02	0.05	0.14	0.13	0.36
.11	-0.24	0.22	0.14	0.35	0.16	-0.15	0.02	0.00	0.02	-0.13	0.06	-0.00	-0.17	0.06	0.16	0.10	-0.08	0.06	0.09	0.07
	0.02	0.01	0.07	0.16	0.21	-0.05	-0.02	-0.03	-0.09	0.03	-0.05	0.21	0.00	-0.14	-0.04	-0.04	0.06	-0.06	-0.06	-0.10
.02	1	0.82	0.01	-0.33	0.36	0.01	0.11	0.22	0.19	0.25	0.24	0.16	0.48	0.07	-0.15	-0.20	0.04	0.25	0.01	-0.22
.01	0.82	1	-0.11	-0.43	0.45	-0.09	0.22	0.12	0.20	0.23	0.25	0.20	0.78	0.25	-0.04	-0.16	0.03	0.22	-0.00	-0.31
.07	0.01	-0.11	1	0.52	-0.01	0.56	0.44	0.41	0.39	0.24	0.74	-0.36	-0.22	0.41	0.40	0.39	0.04	0.31	0.50	0.50
.16	-0.33	-0.43	0.52	1	0.04	-0.00	0.12	0.07	0.01	-0.06	0.14	-0.09	-0.37	0.04	0.25	0.24	0.18	0.15	0.25	-0.02
.21	0.36	0.45	-0.01	0.04	1	-0.47	-0.09	-0.13	-0.15	-0.00	0.11	0.14	0.43	-0.03	-0.07	-0.21	0.18	0.27	-0.07	-0.51
.09	0.01	-0.09	0.56	-0.00	-0.47	1	0.05	0.03	0.04	0.05	0.43	-0.13	-0.17	0.07	0.02	0.01	-0.02	0.03	0.01	0.73
.02	0.11	0.22	0.44	0.12	-0.09	0.05	1	0.46	0.58	0.25	0.56	-0.30	0.07	0.77	0.69	0.59	-0.04	0.02	0.60	0.22
.03	0.23	0.12	0.41	0.07	-0.13	0.03	0.46	1	0.61	0.24	0.41	-0.27	-0.07	0.46	0.36	0.48	-0.15	0.20	0.56	0.13
.09	0.19	0.20	0.39	0.01	-0.15	0.04	0.58	0.61	1	0.50	0.47	-0.31	0.07	0.59	0.57	0.55	-0.20	0.10	0.71	0.24
.03	0.25	0.23	0.24	0.06	-0.00	0.05	0.25	0.24	0.50	1	0.35	-0.25	0.16	0.34	0.29	0.27	-0.05	0.11	0.31	0.06
.05	0.24	0.25	0.74	0.14	0.11	0.43	0.56	0.41	0.47	0.35	1	-0.36	0.13	0.62	0.47	0.33	0.04	0.28	0.48	0.34
.21	0.16	0.20	-0.36	-0.09	0.14	-0.13	-0.30	-0.27	-0.31	-0.25	-0.36	1	0.12	-0.48	-0.41	-0.35	0.16	-0.17	-0.43	-0.21
.00	0.48	0.78	-0.22	-0.37	0.43	-0.17	0.07	-0.07	0.07	0.16	0.13	0.12	1	0.13	-0.07	-0.22	0.05	0.15	-0.11	-0.35
.14	0.07	0.25	0.41	0.04	-0.03	0.07	0.77	0.46	0.59	0.34	0.62	-0.48	0.13	1	0.72	0.60	-0.10	0.12	0.59	0.18
.04	-0.15	-0.04	0.40	0.25	-0.07	0.02	0.69	0.36	0.57	0.29	0.47	-0.41	-0.07	0.72	1	0.76	-0.20	-0.02	0.67	0.30
.04	-0.20	-0.16	0.39	0.24	-0.21	0.01	0.59	0.48	0.55	0.27	0.33	-0.35	-0.22	0.60	0.76	1	-0.21	-0.02	0.65	0.30
.06	0.04	0.03	0.04	0.18	0.18	-0.02	-0.04	-0.15	-0.20	-0.05	0.04	0.16	0.05	-0.10	-0.20	-0.21	1	-0.42	-0.23	-0.27
.06	0.25	0.22	0.31	0.15	0.27	0.03	0.02	0.20	0.10	0.11	0.28	-0.17	0.15	0.12	-0.02	-0.02	-0.42	1	0.00	-0.16
.06	0.01	-0.00	0.50	0.25	-0.07	0.01	0.60	0.56	0.71	0.31	0.48	-0.43	-0.11	0.59	0.67	0.65	-0.23	0.00	1	0.25
.10	-0.22	-0.31	0.50	-0.02	-0.51	0.73	0.22	0.23	0.24	0.06	0.34	-0.21	-0.35	0.18	0.30	0.30	-0.27	-0.16	0.25	1

puede representar todos los aspectos de un país, región, municipio, ciudad, etc, altitud, clima, vegetación, población, actividades productivas, aspectos sociales, opiniones políticas o bien puede considerar un aspecto social, demográfico o económico.

Los elementos de la M.I.E., (Unidades de observación y atributos espaciales), es la información que se perfora en fichas del ordenador e integran los datos de base. Todas las secciones o demarcaciones llevan un número igual de tarjetas y la cantidad de éstas depende del número de variables de la matriz y de la cantidad de dígitos que se use.

y viceversa. El mismo juego se realiza cuando disminuye.

R: 0; Muestra que no hay correlación, que no existe ningún tipo de nexo ni de relación. Toda aproximación a cero gradua la intensidad de la independencia de X_1 y x_2.

R: —1; Significa una dependencia máxima lineal decreciente negativa entre x_1 y x_2. O sea, que cuando aumenta X_1 y X_2 disminuye en un porcentaje similar al aumento de la primera variable y viceversa.

En realidad el índice de correlación lineal mide los lazos aparentes de los que no se deduce de inmediato una relación de casualidad entre las variables. Un alto índice de correlación puede ser debido a:

1.º Existencia de una relación causal entre los fenómenos que representan las dos variables (x_1 x_2). En este caso una variable está causada por la otra.

2.º Los dos atributos actúan uno sobre otro de forma recíproca.

3.º Las dos variables están influidas por una causa común exterior.

4.º Sólo hay una simple concomitancia de variación (covariación); éste caso puede ser identificado mediante posteriores análisis estadísticos.

5.º Las variaciones conjuntas pueden ser debidas al azar, o también ocasionadas por la falta de datos o por su débil fiabilidad (caso de correlación no significativa) (6).

En la práctica para que dos variables (x_1, x_2) tengan una correlación significativa, su coeficiente debe de estar fuera, es decir por encima de + 0,3 o por debajo de — 0,3. Por lo tanto la indiferencia o independencia de x y x se dará cuando está comprendido su índice entre + 0,30 y — 0,30 (7).

Anteriormente apuntábamos que la matriz de correlaciones (M.C.) es cuadrada: R: (m x m), siendo "m" las variables o atributos. En la figura geométrica resultante la diagonal contiene la correlación de la variable consigo misma siendo por lo tanto su valor 1. En la M. C. los coeficientes de correlación se repiten y son idénticos a uno y otro lado de la diagonal. A continuación añadimos un aspecto parcial de una M. C. de un estudio para una mejor comprensión (Cuadro n.º 2).

En la muestra anterior están subrayados los coeficientes de correlación significativos. Estos se obtienen mediante un test de significación en los que el coeficiente de correlación aumenta conforme el número de U.O. disminuye. Si el margen de error se reduce, se incrementa lógicamente el tamaño del coeficiente.

A fin de facilitar la comprensión del contenido teórico de las correlaciones, consideramos pedagógico comentar algunas de las que se han obtenido en esta muestra parcial que intercalamos en el texto.

La variable segunda, la del estado de casado correlaciona positivamente con los solteros (0,84), viudos (0,57), con la población de 20 a 50 años (0,39), nacidos en la localidad (0,39), analfabetos (0,45), adultos con Bachiller Superior y Universidad, etc. Al ser las correlaciones positivas, ésto indica que al aumentar los casados también aumentan los solteros, los viudos, las personas de 20 a 59 años, los nacidos en la localidad y viceversa, etc. Ahora bien, los coeficientes no son iguales, elevados al cuadrado cuantifican el tanto por ciento que explica la variación de una variable por la influencia de otra.

Cuando varían los casados, el 64% de esa varianza está explicada por la influencia de los solteros, el 25% por la de los viudos, el 16% por la de 20 a 59 años, etc. Otro ejemplo podría ser la variable n.º 6, la población de 0 a 19 años que correlaciona con la población de 0 a 4 años (0,82), con los nacidos en la localidad (0,45), estudiantes de primaria (0,78) (y negativamente con la población de 60 años y más (—0,43).

Al incrementarse los niños y jóvenes de 0 a 19 años, aumenta en un mismo conjunto humano considerado los niños de 0 a 4, los nacidos en la localidad y los estudiantes de Primaria y desciende por el contrario el número de viejos de 60 y más años. Elevando el coeficiente de correlación de cada uno, da el porcentaje de dependencia entre ellos (r = 67%) para los niños de 0—4, etc. En el caso negativo al aumentar los individuos de 0 a 19 años en 100, descienden los viejos en un 18,5%. Cosa lógica ya que cuanto más jóven es una población el porcentaje de viejos desciende. En este caso concreto sabemos cual es la relación e influencoa de otras variables en la variación.

Un segundo aspecto que permite iniciar una explicación general de todos los hechos relacionados con la M.C. y mostrar su interacción y sus posibles causas es la técnica denominada "Linkage analysis", desarrollada por L.L. Mc. Quitty (8) para sus estudios en psicología y que fue utilizado posteriormente por los geógrafos tras su introducción en nuestra disciplina por J. B. Berry (9).

Mc. Quitty define como "linkage" para cada variable de una matriz el lazo o coeficiente de correlación más elevado que une a una variable con otra de la misma matriz de correlaciones". Mediante éste análisis realizamos la identificación y el reconocimiento de las llamadas *parejas de variables recíprocas* y que son aquellas que tienen el máximo índice de correlación de una con la otra, tanto positiva como negativa y viceversa. Por ejemplo, el primer par recíproco podría ser el integrado por las variables x_1 y x_2 ; donde el máximo coeficiente de correlación de X_1 con todas se da con la x_2 , y el máximo coeficiente de correlación de x con todas se realiza con la x_1.

$$x_1 \rightleftharpoons x_2 \quad \text{donde} \quad r = 0'841$$

Simultáneamente cada pareja recíproca puede ser el núcleo de una cadena de variables existentes en la M.C. Para construir estas cadenas se parte de las parejas recíprocas detectadas, y en las restantes variables se buscan las que mantienen con alguna de las del par, el índice de correlación más elevado y así sucesivamente hasta completar todas las cadenas. A éstas se les llama "primas" y conforme se alejan del par son "primas segundas", "terceras", etc.

Al final todas las variables están integradas en alguna cadena, que en función del signo del coeficiente son positivas o negativas. El interés de este método de trabajo radica en que no se necesitan grandes medios, no siendo imprescindible el uso del ordenador, si la M.I.E. no es muy grande, ya que los coeficientes de correlación se obtienen de forma menos complicada (10).

Las conclusiones que de las cadenas se extraen son interesantes y profundas, siendo en cierto modo un avance de la estructura que manifestarán posteriormente los factores. Son pues, minifactores que unifican un gran número de variables y que sirven para sintetizar hechos variados y diversos.

En los cuadros nº 3 y 4, inscribiremos la relación de variables con sus correlaciones máximas negativas y positivas y el número de la variable con que se liga, pertenecientes a un estudio tipo.

3. *La Matriz Factorial.*

Es una matriz de orden A = M x P, en donde M son las variables y P son los factores o componentes principales. Son extraídos de esta matriz mediante el criterio clásico de considerar los componentes principales cuyo autovalor "eigenválues" sea igual o superior a uno.

El número de autovalores los dá directamente el programa y van de mayor a menor, nosotros sólo tenemos que contabilizar los exponentes que sean superiores a 1. La cantidad que resultase son él número de factores de utilidad máxima que se han obtenido. Después, sólo hay que contar en la matriz factorial el número de columnas que nos reflejan los autovalores.

Obtenida la matriz factorial reducida A' = (M x P), ("M" son las variables y "P" los factores), hay que comprobar y seleccionar en cada columna del listado que pertenece a un factor, las saturaciones de cada variable en el factor que sean superiores a + 0,40. Descendamos a un ejemplo práctico real con una parte de la M.F. reducida.

CUADRO 3

CORRELACIONES MAXIMAS POSITIVAS : CADENAS POSITIVAS

N.º VARIABLE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
N.º VARIABLE CORRELACIONADA	2	1	1	27	58	59	56	27	6	57	18	13	12	13	7	48	59	19	43	43	- -	- -	43
COEFICIENTE CORRELACION	841	841	704	355	999	902	827	582	452	769	772	612	612	507	748	579	833	726	763	788	- -	- -	828
* CADENA	P_1	P_1	1	5	P_2	P_3	7	5	3	13	5	P_4	P_4	4	7	11	3		5	5	- -	- -	P_5

N.º VARIABLE	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
N.º VARIABLE CORRELACIONADA	42	56	16	43		630	32	- -	30	18	67	67	64	67	64	67	67	65	24	23	18	1	24
COEFICIENTE CORRELACION	820	860	462	538		970	499	- -	499	615	871	952	583	913	841	983	967	750	820	828	693	446	719
* CADENA	P_6	P_7	11	5		P_{15}	P_8	- -	P_8	5	10	10	9	10	P_9	P_{10}	10	10	P_6	P_5	5	1	6

N.º VARIABLE	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70
N.º VARIABLE CORRELACIONADA	32	66	12	52	52	51	- -	- -	59	25	61	5	6	70	57	43	29	38	67	48	39	43	- -	60
COEFICIENTE CORRELACION	452	671	401	409	727	727	- -	- -	731	860	791	999	902	612	791	677	970	841	971	671	983	460	- -	612
* CADENA	8	P_{11}	4	12	P_{12}	P_{12}	- -	- -	3	P_7	P_{13}	P_{12}	P_{13}	P_{14}	P_{13}	5	P_{15}	P_9	10	P_{11}	P_{10}	5	- -	P_{14}

NOTAS

* EN LA ULTIMA FILA DENOMINADA CADENA
  – El número: Indica la cadena a la que pertenece la variable
  – P: Variable que forma par reciproco con otra del mismo número: $P_1 \leftrightarrows P_1$; $P_2 \leftrightarrows P_2$

CUADRO 4

CORRELACIONES MAXIMAS NEGATIVAS : CADENAS NEGATIVAS

N.º VARIABLE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
N.º VARIABLE CORRELACIONADA	- -	- -	- -	- -	- -	8	70	6	10	60	66	60	60	- -	70	18	- -	66	66	66	22	21	66
COEFICIENTE	- -	- -	- -	- -	- -	431	716	431	476	742	523	442	502	- -	556	486	- -	643	589	522	425	425	558
CADENA						P_1	5	P_1	5	5	2	5	5	- -	5	P_2	- -	P_2	2	2	P_3	P_3	2

N.º VARIABLE	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
N.º VARIABLE CORRELACIONADA	60	70	62	- -	- -	- -	- -	- -	19	- -	- -	- -	- -	- -	23	- -	- -	- -	60	48	48	- -	- -
COEFICIENTE	651	586	409	- -	- -	- -	- -	- -	425	- -	- -	- -	- -	- -	554	- -	- -	- -	573	570	540	- -	- -
CADENA	5	5	4						2						2				5	P_4	4		

N.º VARIABLE	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70
N.º VARIABLE CORRELACIONADA	20	43	- -	- -	- -	- -	- -	- -	- -	70	60	- -	- -	61	60	48	- -	43	- -	18	- -	25	- -	61
COEFICIENTE	509	570	- -	- -	- -	- -	- -	- -	- -	576	791	- -	- -	999	999	559	- -	423	- -	643	- -	466	- -	612
CADENA	2	P_4								5	5			P_5	P_5	4		4		P_2		5		5

Las saturaciones de cada variable subrayadas en cada factor son las significativas por ser superiores todas a + 0,40. Todas las superiores a ese valor son las que integran y definen el factor y la saturación mayor o menor indica la cuantía de la intervención o influencia en el factor.

Como las saturaciones pueden ser positivas o negativas cada factor puede tener dos polos que son dialécticamente opuestos. Es decir, que cuando se dan o predominan las variables o atributos de un polo las del opuesto no tienen valor o no existen o viceversa. Pero ambos polos integran y definen al factor, forman una estructura dialéctica, no son independientes, son como el anverso y reverso de una moneda, que solos no son nada pero unidos, aunque opuestos, hacen un todo.

Los factores resumen la información, la organizan, cuantifican las correlaciones y las dependencias e implicaciones de las variables en cada factor. Según sea el autovalor del factor explica un porcentaje mayor o menor de toda la varianza general de la información, cuanto más grande es esa explicación porcentual, el resultado en su vertiente positiva se eleva. La mecanica a seguir tiene que partir de la observación de los listados. Aplicando el criterio de los autovalores, obtenemos en el estudio citado un número de 14 componentes que cumplen con este requisito.

Con esos componentes formamos la matriz factorial reducida. Los componentes identificados explican un porcentaje del total de la varianza, cifra que conforme es más elevada englosa un total de información mayor.

No obstante y para ser más operativos, conviene hacer de nuevo una segunda selección subjetiva de los componentes principales matemáticamente obtenidas (eigenvalues), ya que la capacidad de síntesis y de explicación de ellos es muy diferente. De la serie obtenida y de una forma gradual del primero al último va disminuyendo su valor explicativo y por tanto su interés, de ahí que después haya que explicar un criterio selectivo, mediante el cual nos quedamos tan sólo con los componentes cabezas de serie más significativos, los cuales explicaban entre sí un porcentaje alto del total general de la varianza, siendo de fácil denominación por el valor de las saturaciones y las variables contenidas, (Cuadro n.º 5).

Despreciándose los restantes que suelen ser —a veces— una reiteración o repetición degradada de los primeros. En el ejemplo al que aludiamos, el primer factor mantiene coeficientes de saturación significativos (mayores + 0,40) con 26 variables y explica por sí solo un 19,03 por ciento de la varianza original.

El segundo alcanza 14 coeficientes de saturación y explica individualmente el 14,5%.

El tercero, 13 coeficientes de saturación y 9,6 %.

El cuarto, 12 coeficientes de saturación y 8,4 por ciento y quinto, 6 saturaciones superiores a + 0,40 y explica el 5,3 por ciento. Cada factor puede representar varias variables. Los factores disminuyen en importancia y peso de modo que el número Iº es el más importante, el IIº es el siguiente, etc.

El primer factor lleva aproximadamente el 19,03% de la varianza total entre todas las 70 variables, el factor II, 14,5%. En total los seis primeros contienen el 56,8% de la varianza total. El grado de relación de cada variable con cada factor depende su distancia de 0. Así por ejemplo la variable nº 7 (poblac. de 20 — 59 años) tiene — 0,74, con el factor I. La nº 5 y 6 están ligadas con el factor III por — 0,806 y — 0,807. Mientras que las variables n.º 19 (Varones activos) y la 41 (Trabajadores de las finanzas, seguros, etc), respecto al facros I apenas estan ligadas por su bajo coeficiente de saturación: 0,382, y 0,266. En el factor III la variable 15 y 25 no tienen casi valor y significa-

CUADRO 5

FACTORES (1)

	1	2	3	4	5
Autovalores	10.322	10,222	6,750	5,800	3,751
Porcentaje Explicación	19,03 o/o	14,57 o/o	9,60 o/o	8,40 o/o	5,30 o/o
o/o Acumulado	19,03	33,6	43,2	51,6	56,9

N.º Variable	Texto Variable	Saturaciones de las Variables en cada factor				
		1	2	3	4	5
1	Solteros	-.252	.073	-.325	-.073	.240
2	Casados	-.371	.181	-.240	.317	.326
3	Viudos	-.120	-.051	.135	-.197	.536
4	Separados	.113	.056	.051	-.104	.357
5	0 - 4	.089	.077	-.806	.011	-.071
6	0 - 19	.134	.123	-.807	-.188	-.188
7	20 - 59	-.741	.264	.069	.202	.440
8	60 y +	.269	.072	.441	-.249	.534
9	Nac. locales	.331	-.016	-.478	-.369	.522
10	Nac. Provin.	-.417	.315	-.008	.768	-.042
11	Nac. Andl. Orient.	-.690	.059	-.232	-.319	-.136
12	Nac. Andal. Occid.	-.572	.006	-.202	-.165	-.086
13	Nacid. Resto España	-.686	-.014	.253	.248	-.179
14	Alfabetos	-.701	.242	.422	-.082	.209
15	Analfabetos	.565	.080	.012	-.019	-.024
16	Est. Bach. Super.	-.745	-.017	-.054	.427	-.085
17	Estud. Univers.	-.709	-.114	-.160	.381	-.127
18	Adult. Bach. Element.	-.770	-.059	.036	.343	.003
19	Varones Act.	-.391	.406	.386	-.604	.267
20	Varones Parados	.518	-.052	.096	.005	.001
21	Mujeres Activas	-.348	.036	-.023	.411	.499
22	Mujeres Paradas	.040	-.057	.158	.001	.124
23	Trabajad. Agricul.	.123	.795	-.107	.154	-.188
24	Trab. Ind. Aliment.	.106	.475	.311	-.016	.239
25	Trab. Ind. Papel	-.162	-.069	.006	.091	-.020
26	Trab. Ind. Quimicas	.057	.871	-.168	.138	-.109
.						
.						
70						

(1). Es solo una parte de una Matriz Factorial más amplia.

ción en el factor por tener unas saturaciones inferiores a 0,400. El medio de denominar a los factores identificados es un problema final y definitivo del proceso de A. F. En realidad es aquí donde intervienen más básicamente la experiencia y los conocimientos geográficos del investigador.

El nombre debe ser determinado de acuerdo al tipo de variables ligadas al factor y a sus relaciones entre sí. Especialmente se debe considerar si el factor muestra una oposición entre variables, si existen variables ligadas positivamente al factor y otras relacionadas en forma negativa. También se debe ponderar y valorar más en la selección de un nombre apropiado a las variables que se ligan con un coeficiente máximo, ya sea positiva o negativamente. Desde luego es preciso cuidar nombre o denominaciones especialmente mecanicistas que oscurezcan más que aclaren el sentido recogido por los factores.

Un factor puede ser denominado como "Relieve agreste" por ser las variables de una mayor saturación: Elevada altitud, acuciadas pendientes, poco suelo cultivado materiales calizos, fuerte erosión, etc...Otro denominado "Status socieconómicos" lo podrían definir primordialmente las variables de: Altos niveles de instrucción de la población adulta, elevadas categorías profesionales, porcentajes de población estudiantil, viviendas de gran tamaño, en propiedad y con buenas instalaciones, altos niveles de ingresos, etc...

4. *Contenido y significación de los factores.*

El acto de dar nombre implica una tarea de sintetizar todos los atributos especiales, así como los fenómenos ligados que se dan en el factor, en unas pocas palabras balance final que semánticamente equivale al contenido disperso o desglosado en los atributos identificados con el factor.

Esta tarea puede ser facilitada mediante un procedimiento matemático, llamado "rotación de ejes". Parte del hecho de que los factores se configuran espacialmente de forma octogonal y lo que se hace es ajustar las saturaciones de modo que la ortogonalidad sea mayor, perdiendo valor las menos importantes y ganando las demás. El resultado final consiste en obtener un anueva matriz factorial reducida en la cual cada componente principal haya maximizado su coeficiente de saturación con una o dos variables solamente, de este modo el problema del nombre se simplifica. El primer factor detecta la presencia de variables económicas y sociales. El segundo laborales — productivas, el tercero de cronología demográfica.

Hay varios procedimientos de rotación, nosotros hemos usado la "rotación varimax", incluso en el programa B.M.D. — 03D. Este tipo de rotación sólo ajusta los dos primeros factores, y en ellos hemos incluido las saturaciones obtenidas de forma normal y rotadas.

Tras poner los nombres a los factores, y para culminar las fases de A. F., se realiza la obtención de los "pesos o scores" , de los factores en cada unidad de observación de crucial importancia para el geógrafo ya que permite cartografiar las variaciones cuantitativas y espaciales de cada factor.

Los pesos o scores, son los valores que alcanzan cada factor en las unidades de observación. Los pesos adquieren por tanto unos valores positivos o negativos según predomine en esa unidad espacial las variables de un polo u otro. Si el peso de un factor en una U.O. es positivo indica que los atributos o variables predominantes son las del polo positivo, si el peso es negativo esa U.O. estará denominada por las características del polo negativo del factor. El ordenador da finalmente una matriz de N x S, donde N son las unidades de observación y S, son los pesos de cada factor, con valores positivos o negativos. (Cuadro nº 6). Para obtener una más objetiva representatividad cartográfica

CUADRO 6

PESOS INDIVIDUALES DE LOS CINCO PRIMEROS FACTORES

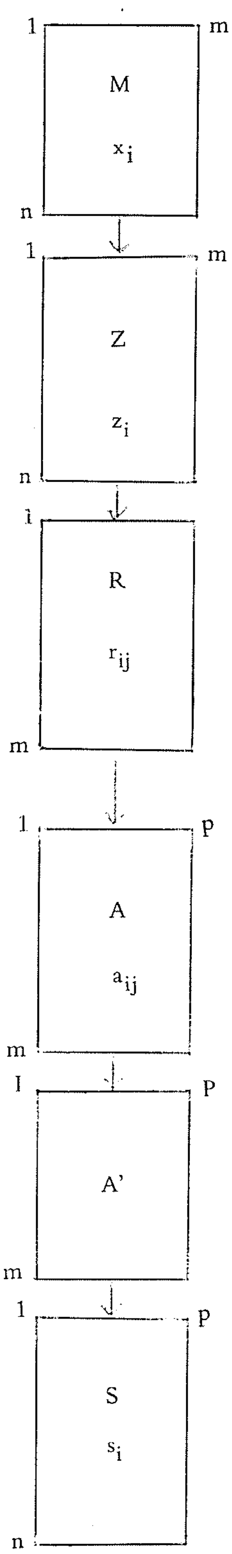
FACTORES (1)

Secciones Urbanas	1	2	3	4	5
I- 1	-7.81	-0.13	9.03	-4.04	2.73
I- 2	0.64	-0.76	6.22	0.87	6.21
I- 3	-10.52	-3.63	10.18	-1.77	4.76
I- 4	-10.41	-3.74	10.02	-4.50	2.09
I- 5	-8.51	-4.63	9.84	-2.35	-0.96
I- 6	-8.37	-1.10	3.99	-2.74	1.88
I- 7	-9.67	-0.71	1.82	-4.06	-5.29
I- 8	-10.89	-1.04	-3.42	-5.80	-4.60
I- 9	-12.73	-1.20	1.53	-5.49	-1.36
I- 10	-8.23	1.07	0.22	-3.50	-3.01
I- 11	-6.74	-0.53	3.08	-3.10	-5.51
I- 12	-8.14	1.81	-5.43	-1.79	-3.18
I- 13	-2.20	3.59	-8.45	2.48	1.17
I- 14	-20.89	3.10	-8.79	4.22	-1.14
II- 1	-0.20	-3.07	7.53	-1.41	2.60
II- 2	-10.19	-4.79	9.93	-4.02	-0.62
II- 3	-6.45	-1.63	8.72	0.68	2.89
II- 4	-9.64	-3.07	6.52	-3.93	3.37
II- 5	-6.11	-1.92	4.78	-2.70	10.37
II- 6	-5.78	-2.25	3.28	0.15	0.37
II- 7	-15.35	-2.72	-2.48	-8.35	-3.24
II- 8	-15.21	-2.35	2.87	-5.45	-2.69
II- 9	-10.30	-0.34	-1.10	0.12	0.19
II- 10	-7.57	-0.65	-0.13	-2.11	-3.92
II- 11	-10.68	-2.19	4.58	-4.03	1.76
II- 12	-11.63	-3.74	1.62	-6.81	-4.61
II- 13	-10.90	0.77	-11.00	-0.18	1.77
II- 14	-10.38	0.00	-1.95	-0.92	-1.31
II- 15	2.28	-2.01	0.14	3.73	2.00
II- 16	11.97	-1.98	-0.81	0.72	-3.26
II- 17	6.58	-1.69	-4.10	2.72	1.62
II- 18	6.43	-0.26	-2.43	2.36	1.81
II- 19	2.99	0.79	-8.88	2.97	-1.08
II- 20	-6.79	0.09	-4.50	0.83	9.75
II- 21	-0.57	1.25	0.87	2.59	5.07
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
N.º 120					

(1). Es solo una parte de una matríz más amplia de pesos o "scores".

GRAFICO 1

LAS ETAPAS DEL ANALISIS FACTORIAL



1. MATRIZ DE INFORMACION ESPACIAL ORIGINAL
(N observaciones) x (M variables)
x_i = Valor general de los elementos de esta matriz

2. MATRIZ DE INFORMACION ESPACIAL TIPIFICADA
(N observaciones) x (M variables)
z_i = Valor general de los elementos de esta matriz obtenido por la fórmula. $z_i = \frac{x_i - m_i}{i}$

m_i = Valor medio general de las variables
i = Desviación típica general de las variables

3. MATRIZ DE CORRELACCIONES ENTRE LAS VARIABLES
(M variables) x (M variables)
r_{ij} = Valor general de los elementos de esta matriz.

$$r_{ij} = \frac{\Sigma (x - \bar{x}) \cdot (y - \bar{y})}{\Sigma (x_i - \bar{x}_i) \cdot (y_i - \bar{y}_i)^2}$$

$\bar{x}$ = Valor medio general de una variable
$\bar{y}$ = Valor medio general de otra variable.

4. MATRIZ FACTORIAL
(M variables) x (P factores)
a_{ij} = Valor general de los elementos de esta matriz
$x_i = \sum^{p} a_{ij} F_j$, y se cumple que la matriz de correlaciones es igual a la matriz de saturaciones (a_{ij}) ó (A) multiplicada por su transpuesta.

$(R) = (A) \cdot (A)^T$ ó $(rij) = (aij) \cdot (aij)^T$

5. MATRIZ FACTORIAL SIMPLIFICADA (ROTACION VARIMAX)
(M atributos espaciales) x (P componentes principales)

6. MATRIZ DE INFORMACION ESPACIAL REDUCIDA. Peso local en cada factor de cada observación.
(N observaciones) x (P factores)
Se obtiene multiplicando la matriz A por la Z

$(S) = (A) \cdot (Z)$ ó $(s_i) = (a_{ij}) \cdot (z_i)$

Al término de todo el proceso anterior obtenemos una nueva matriz de información espacial pero ahora con muchos menos elementos y conservando, sin embargo, la mayoría de la información inicial contenida en nuestra primera matriz. A partir de esta nueva matriz se puede continuar el análisis, sabiendo que ahora partimos de datos que sintetizan los datos de base, y en realidad se puede considerar que los factores determinados son hechos estructurales que se expresan a través de los datos originales.

se calcula la media de todos los pesos por cada factor y la desviación estandar. Con ambos índices de centralidad y dispersión se establece una tabla de frecuencias y una trama para cada intervalo que se pasa a un plano. La cartografía factorial ayuda a entender el verdadero significado de los factores determinados.

Una última opción, muy interesante y útil para la Geografía es la confección de una matriz de correlaciones R = (N x N); donde N son las unidades de observación. Mediante estos coeficientes podemos efectuar un "Linkage analysis" en donde quedaran evidenciadas las zonas o espacios de máxima homogeneidad, homogeneidad media, mínima e indiferencia. Las correlaciones negativas implicarían antagonismo y heterogeneidad. De ello podemos extraer una división territorial en bases objetivas matemáticamente demostrables: divisiones regionales, municipales, barrios o conjuntos urbanos definidos por características y atributos semejantes.

5. *Las depuraciones en el A.F.*

En los AA.FF. simples y más concretamente en los que tienen que efectuarse en sucesivos rangos por necesidades derivadas de contar con un bloque de variables informativas excesivamente grande hay que realizar toda una serie de depuraciones de los resultados obtenidos en cada fase del proceso.

En el caso segundo, se procede a aplicar el A.F. a las diversas tramas concluyendo con un A.F. último sumatorio, en el que se toman como variables los "pesos" de los factores o las variables más relacionadas con los factores obtenidos en cada A.F. de las tramas, teniendo en cuenta los valores más elevados de sus saturaciones, tanto positivas como negativas (ver gráfico nº 2).

La busqueda y aplicación de unas condiciones más óptimas de experimentación implica un desarrollo más largo y complejo de la técnica. Pero ciertamente el ir separando la ganga de la mena en cada fase del proceso redunda en una mejor y más solvente obtención de conclusiones. Esto, hay obligatoriamente que realizarlo, sobre todo, en el caso de matrices de información muy amplias, ya que en ellas una gran parte de los atributos espaciales no alcanzan ni correlaciones ni saturaciones significativas, pero por el contrario, la acción de estas "*variables muertas*" sobre los "pesos" finales puede llegar a ser relativamente importante sesgando y confundiendo los "scores" de la matriz reducida.

En cualquier caso, siempre son elementos de error que se van arrastrando de fase en fase y cuya incidencia malean los "pesos". La solución que se impone consiste en eliminar tras un primer A.F. las variables poco ligadas al conjunto del sistema y realizar con la información depurada un segundo A.F. que será el que sirva de base al cálculo de los "pesos". Como se puede comprobar esto exige una —cuanto menos— duplicidad en el uso y manejo del ordenador para cada A.F. (gráfico nº 3).

Las bases para eliminar las variables poco ligadas se hacen en razón de formular las hipótesis de :

a) Que cada atributo o variable incluida en la M.I.E. tiene un papel en la definición de la estructura del sistema. Esto se mide con las *comunalidades* de cada variable.

b) Que cada atributo considerado tiene relaciones de colinealidad con uno al menos de los restantes atributos. Este se mide analizando la matriz de coeficientes de correlación entre los atributos.

De este modo un atributo que no cumpla al menos una de las dos hipótesis debe eliminarse del análisis. Esto sólo es posible efectuarlo una vez realizado el primer A.F. que nos dará las comunalidades de cada variable. Por lo tanto esto requiere AA.FF. de distintos rangos. Para medir la colinealidad de cada variable se usa el test de Fisher—Student, que según el número

GRAFICO 2

ANALISIS FACTORIAL DEPURADO

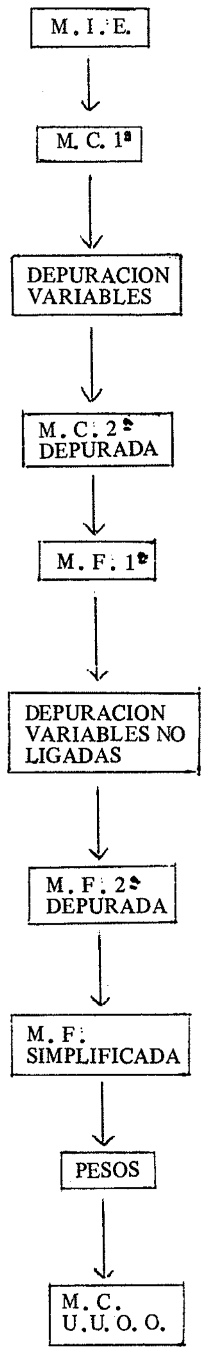
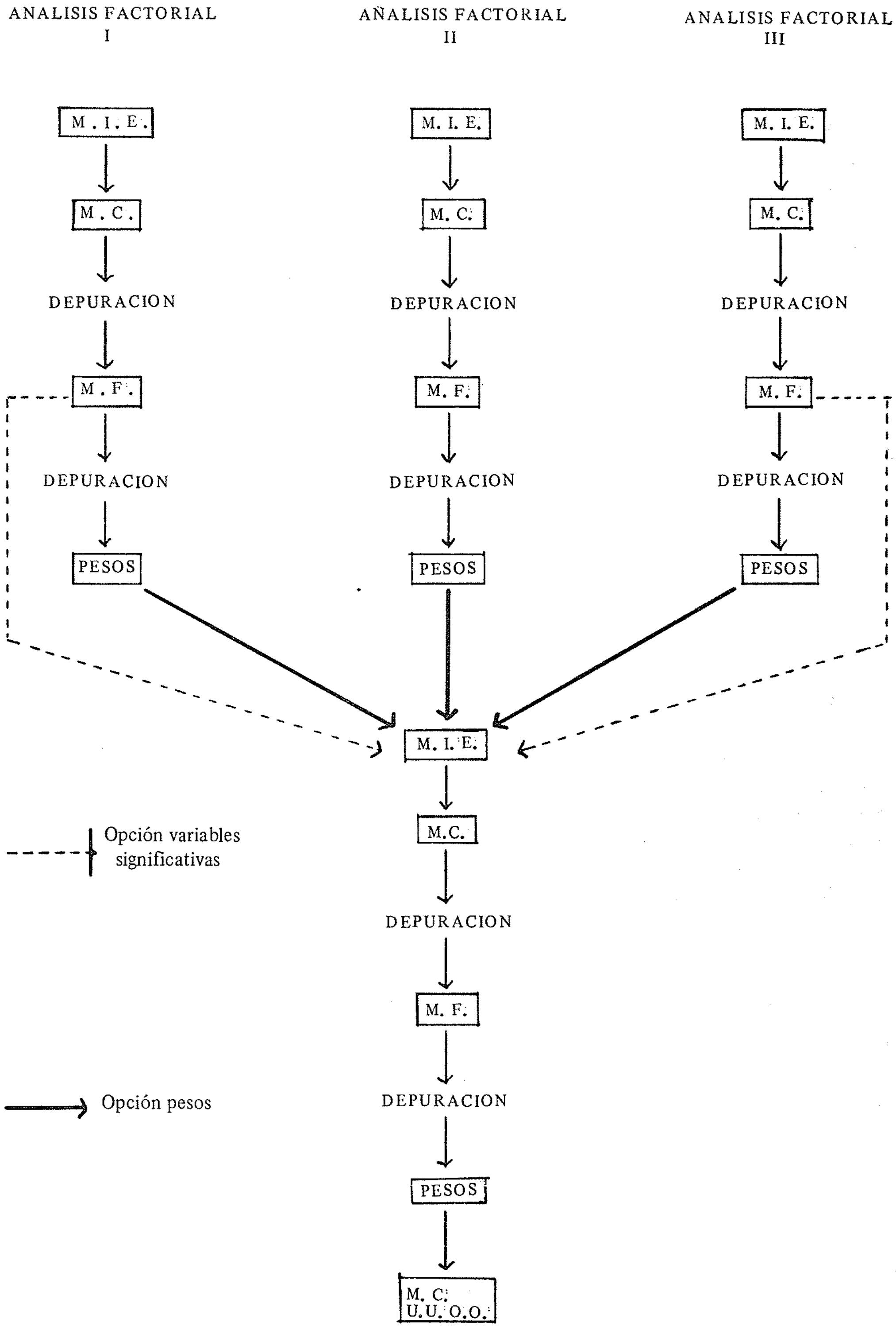


GRAFICO 3

ANALISIS FACTORIALES POR TRAMAS Y RANGOS



de observaciones de cálculo de las correlaciones proporciona un umbral para el coeficiente de correlación con un porcentaje de posibilidad muy alto de que la correlación sea significativa. Los coeficientes de correlación que estén por debajo de éste umbral deben ser desechados.

Un sucedáneo de éste largo proceso de depuración supeditado en gran parte al uso frecuente del ordenador consistiría en recurrir a otros indicadores simples con una pérdida mayor de fidelidad e información general. Existen varias posibilidades en este sentido:

a) Se sacan los atributos o variables que tengan la máxima saturación en cada factor, partiendo del hecho de que puede representarlo con bastante fidelidad.

Si se analiza la correlación existente entre los "pesos" locales del factor y valores del atributo de cabeza en cada observación, podemos determinar si es aceptable y cuanta información se pierde. Racine lo hace y, en su caso, sólo se le resta el 10 por 100 de la información.

b) También se puede considerar para elegir el atributo representativo, lo que conceptualmente represente el factor. Con estos criterios lo que se hace preferentemente es elegir las variables de mayor saturación y más significado en cada factor y usarlas en AA.FF. de segundo y tercer rango, desechando subjetivamente el resto de los atributos. Con ellos se puede proceder al A.F. definitivo sobre estos atributos o variables representativos del conjunto o indicadores compuestos.

III. CONCLUSIONES METODOLOGICAS.

1.ª Aunque la aplicación de la técnica del A.F. es extraordinariamente útil realizada individual y aisladamente como recientemente se ha venido demostrando en los estudios geográficos, nuestra experiencia nos ha venido igualmente a demostrar que realizada con posterioridad a un análisis tradicional descriptivo y causal del objeto en estudio, posibilita unos mejores éxitos en cuanto a su aplicación, matización y profundidad de sus resultados.

En resumen un conocimiento previo de la realidad, un estudio descriptivo favorece una mejor interpretación y mayor riqueza de conclusiones del A.F.

2.ª No hay que pensar en el A.F. como una panacea que ofrece soluciones ante una compleja realidad. Es un instrumento de investigación matemático —estadístico complejo y difícil de manejar si se le quiere sacar rendimiento. Como todo proceso de investigación, requiere unas condiciones técnicas de trabajo idoneas, así como una delicada y continua depuración en sus sucesivas fases. Es necesario ir aislando lo esencial y desechando lo que la misma técnica nos señala como superfluo y accesorio.

3.ª Es una técnica válida y compatible con la formulación de hipótesis geográficas iniciales, reconociendo su validez de comprobante y su capacidad de demostración empírica.

NOTAS BIBLIOGRAFICAS

1. *Sobre Análisis Factorial.*

— RACINE, J.B. et RAYMOND, H. "L'analysi quantitative en Geographie" P.U.F. Paris 1973.

— NAYRAC, J. "Componentes et facteurs, méthodes pour le dégagement des concepts generaux en psychologie quantitative". Paris 1961.

— HERMAN, H.H. "Modern Factor Analysis" Chicago 1967. University of Chicago Press, 2ª edic.

— DENIUU, C. CEBART, L. "Introduction à l'analysis des donées" in Consommation CREDOC. 1967 . 3 . 57 — 96 y 4 . 65 — 87.

— BERRY, B.J.L. "DIDO DATA analysis" 160 or PATTERN recognition?. Perspectives in Geography, Models of spatial variation. Northen Illinois Univerity, 1971.

— BERRY, B.J.L. "City classification handbook: methods and applications" Wiley Interscience. New York, 1972.

— KING, L.J. "Cross-Sectional analysis of Canadian Urban Dimension" 1951 an 1961. Canadian Geographer. Vol. 10, nº 4, pp. 205—224.

— SOLA-MORALES, M. "Factorización de características de un área suburbana" Rev. de Geografía de la Universidad de

— THURSTONE, L.L. "Múltiple factor analysis" Universidad Chicago Pres 1947.

— TORRENS—IBERN, J. "Modeles et Methodes de l'Analysis Factorielle" Dunod. Paris 1972.

2. RACINE, J.B. "Ecología factorial y ecosistemas espaciales" en "Perspectivas en Ecología Humana de BOURGOIGNIE, G.E. Instituto de Estudios de la Administración Local. Madrid 1976, p. 20.

3. D.O. PRICE: "Factor Analysis nithe study of Metropolitan Centers" Social Forces, 20, 1941—42, pp. 449 y ss.

4. Programas más usuales. De éstos programas los más importantes y utilizados son:

— EL FACTO, muy simple, no ofrece opciones para el investigador. Cuenta con dos limitaciones: sólo se pueden utilizar veinte y nueve variables y 30.000 observaciones.

— EL FCTR, más complejo. Ofrece la opción de introducir, como dato original, la matriz de datos o la matriz de correlaciones. Además, es posible realizar rotaciones de los factores, rotaciones oblicuas. Estas rotaciones consisten en una técnica que acerca y maximiza el valor de la saturación de una variable con un factor y minimiza el resto; de este modo el problema de identificar los factores se reduce. Limitaciones: sólo se puede utilizar este programa con 499 observaciones y 30 variables. Estos dos programas son para ordenadores IBM.

— Para los ordenadores UNIVAC existen los programas: el BMD 03M, ofrece gran número de opciones, incluso generar variables a partir de las variables originales. Realiza la rotación de ejes o factores "Varimax", rotación que mantiene la perpendicularidad de los factores. Sus limitaciones son: sólo 80 variables y 9999 observaciones.

— Por último, la subrutina FACTAN, programa muy parecido al FACTO pero adaptado al ordenador UNIVAC; este programa obtiene: la matriz de correlaciones, la matriz factorial sin rotar y la matriz de peso de las observaciones en cada factor.

5. RACINE, J.B. et REYMOND: "L'analyse quantitative..." opus cit, p. 131.

6. Ibidem p. 135.

7. Un índice de correlación de + 0,30 nos dá un coeficiente de determinación de r = 9 %. Menos del 10 % de la variación de x se explica por influencia de x de x , lo cual es poco significativo.

8. L.L.Mc. QUITTY. "Elementaru Linkage Analysis for Isolating Orthogonal and oblique types and typal Relevancies" Educational and Psychological Measurement, Vol 17, 1957, pp. 207—229. Citado por RACINE et RAYMOND, en "L'analysis..." opus cit, p. 151.

9. BERRY, B.J.L. "Shopping Centers and the Geography of Urban Areas" A Thoerical and Empirical Study of the Spatial Structure —of urban—. Retail and services Busi-ners. University of Washingtons, Ph. Dr. Thesis 1958.

10. $$rij = \frac{\Sigma(x_i-\bar{x})\,(y_i-\bar{y})}{\sqrt{\Sigma(x_i-\bar{x})^2\,(y_i-\bar{y})^2}}$$

$\bar{x}$ = Valor medio de una variable
$\bar{y}$ = Valor medio general de otra variable.

11. Los pesos se obtienen multiplicando la matriz de información tipificada por la matriz factorial (factores elegidos). El resultado es una matriz reducida "S": (N. observaciones) x (P: componentes principales).

12. RACINE, J.B., CAVALIER, M. "Ecologie factorielle et atributs geographiques" Cahiers de Geographie de Quebec, 1972, p. 213.

13. RACINE, J.B. "La croissance du Grand Montreal: Geographie factorielle d'un phénomene suburbain" Ottawa, 1972 Edition de l'Université d'Ottawa. Thése de doctorat d'Etat y RACINE, J.B. y CAVALIER, M. "Ecologie factorielle et atributs geographiques" Cahiers de Geographie de Quebec, 1972.