

## EL FUTURO DE LA GEOGRAFIA CUANTITATIVA

### INTRODUCCION

En los últimos diez años se ha producido una "revolución" en la revolución cuantitativa. A comienzos de la década de los años setenta, los geógrafos cuantitativos han dirigido la atención desde los complejos modelos multivariados y las técnicas estadísticas clásicas que fueron tan utilizadas en los años 60, hacia una amplia gama de nuevas técnicas más adecuadas al tratamiento de problemas de investigación actual y a los tipos de datos con que cuentan los geógrafos.

El desarrollo pasado en la geografía cuantitativa reflejaba también un desarrollo global de toda la geografía. De esta manera, al comenzar la revolución cuantitativa a principios de la década de los años 50, el interés y la difusión fué general y así se aplicaron índices, coeficientes de localización, modelos de correlación y regresión simple a todas las ramas de la geografía humana. En los años 60 el gran desarrollo de la geografía urbana, como lo atestiguan los numerosos trabajos de Brian Berry, produjo un rápido incremento en el uso de técnicas estadísticas multivariadas, y de un modo especial el análisis de componentes principales, análisis factorial y análisis de agrupamiento. El denominador común de estos estudios fue el intento de comprender la estructura interna de la ciudad mediante el empleo de conceptos y técnicas de ecología factorial.

Al comenzar la década de los años setenta, la ecología factorial fué objeto de ataques desde posiciones diferentes. Los geógrafos radicales norteamericanos señalaban que a pesar del gran número de ecologías factoriales producidas nada

se había hecho en atenuar la miseria urbana. Teóricos tales como David Harvey manifestaban que el enfoque ecológico fue usado frecuentemente como un sustituto, más que como una *ayuda* para desarrollar un marco teórico dentro de la geografía urbana y que la preocupación aparecía más en conseguir resultados de tipo académico que en comprender el funcionamiento de la ciudad.

Por otra parte, la ecología factorial fué objeto de críticas de los mismos geógrafos cuantitativos que pusieron de manifiesto la necesidad de resolver muchos problemas técnicos antes de aplicar de nuevo el método factorial al estudio de la estructura urbana. Estos problemas incluyen: la selección de un modelo de análisis factorial apropiado; el problema de las unidades espaciales objeto de estudio; la influencia de la autocorrelación espacial en los datos analizados; etc. Un buen ejemplo de estos planteamientos críticos abordado por geógrafos cuantitativos puede verse en GIGGS, J.A. y MATHER, P.M. (1975).

A comienzos de la década de los setenta muchos de los conceptos neopositivistas que constituían los pilares de la geografía cuantitativa, son objeto de un análisis crítico severo. El rechazo de las ideas más simplistas sobre la concepción de la *geografía como una ciencia espacial* condujo a introducirse en campos de estudio más amplios tales como los propuestos por la geografía del comportamiento y radical. Sin embargo, finalizada la década de los años setenta estas nuevas preocupaciones se fragmentaron en un gran número de campos de interés más específico tales como marxismo, anarquismo, feno-

menología, estructuralismo, realismo, humanismo, etc. Esta explosión de nuevos enfoques fué acompañada de un cambio profundo en la geografía cuantitativa. En efecto la aplicación un tanto simplista de técnicas multivariadas, tan características de los últimos años de la década del sesenta, fué sustituida por un interés más amplio en una serie de temas tales como análisis de series temporales, modelos espaciales, análisis de datos categóricos etc. A finales de los años setenta estos temas se subdividen en áreas de estudio mucho más específicas como series espacio-temporales, el uso de modelos *probit* y *logit* con datos categóricos, etc.

El desarrollo de la geografía cuantitativa en el último decenio ha sido tan grande que ya resulta imposible a cualquier geógrafo estar al día en todos los campos y temas en la geografía cuantitativa. La especialización creciente es inevitable y origina el peligro de fragmentación y de duplicación de esfuerzos. A fin de contrarrestar estos peligros en los Estados Unidos y en el Reino Unido, se están haciendo grandes esfuerzos tendentes a la coordinación de las investigaciones, a la difusión y publicación de resultados y ensayos pioneros, como veremos más adelante.

#### PRINCIPALES RESULTADOS EN GEOGRAFÍA CUANTITATIVA DESDE 1.970

No es posible en un trabajo de esta naturaleza, dar una visión completa de todos los resultados y desarrollos producidos dentro de la geografía cuantitativa en el último decenio, no obstante los interesados en un análisis más exhaustivo pueden consultar los trabajos excelentes de Unwin, D.J. (1978), Wrigley, N. y Bennett, R.J. (1981), así como los cuadernos *Catmog* (Concepts and Techniques in Modern Geography) publicados por el Quantitative Methods Study Group (QMSG), adscrito al *Instituto de Geógrafos Británicos* (I.B.G.). (1).

Nosotros pretendemos hacer una breve síntesis de la investigación con el objetivo de proporcionar indicaciones sobre una variada gama de temas abordados en geografía cuantitativa así como señalar las tendencias que se apuntan en este campo de investigación. Aunque por razones de exposición presentamos los resultados bajo apartados diferentes, es preciso recordar que

estos epígrafes se refieren más bien a grupos temáticos que a campos de investigación independientes y por lo tanto, aparecen muy a menudo solapamientos entre los diferentes grupos temáticos.

#### Análisis estadístico multivariado

Como ya indicábamos, el estudio de técnicas multivariadas como objeto de investigación es, en el momento presente, mucho menos frecuente entre los geógrafos cuantitativos. Sin embargo, técnicas tales como análisis factorial y de componentes principales, así como análisis de agrupamiento, se siguen utilizando con gran profusión en muchos trabajos empíricos y en tesis doctorales.

Aunque menos populares que en el pasado, aún se continua en la investigación de estas técnicas, para hacerlas más aptas y apropiadas a los problemas de carácter geográfico. En este terreno se subraya cada vez más la aplicación práctica de estas técnicas. Así por ejemplo, Elffers ha defendido recientemente el uso del "Guttman Criterion Value" como guía para establecer el grado de indeterminación de factores obtenidos en el análisis factorial, (Elffers, H. (1980)). Sin detenernos a comentar las dificultades de aplicar la solución propuesta por este autor, ilustra sin embargo, la creciente preocupación por el desarrollo de técnicas que ayudan a interpretar resultados.

Otro aspecto que ha sido objeto de especial atención fué el de las bases filosóficas en las que se apoyan las técnicas multivariadas. En este sentido, merece especial mención los trabajos de Berry, B.J.L. (1971) y Mather, P.M. (1981), sobre las relaciones entre el análisis factorial y el método científico, relación que no siempre aparece nítida, aunque puede ser fundamental a la hora de aplicar una técnica.

También es preciso destacar, la adopción de estrategias y procedimientos que permiten el empleo de datos no cuantitativos (binarios, nominales y ordinales). Una revisión crítica de este problema aparece en Mather, P.M. (1976) y Anderberg, M.R. (1973). Mención especial merece el trabajo sobre análisis de agrupamiento de Gower, J.C. (1971), que ideó un coeficiente general de disimilitud (*Sg*) que puede aplicarse a un con-

junto de datos mixtos. El coeficiente tiene pesos diferentes según se aplique a datos binarios, categóricos y cuantitativos.

Por último, es de destacar la tendencia al empleo de técnicas multivariadas diferentes a las utilizadas tradicionalmente en geografía. Así por ejemplo, P. Mather en su conocido texto sobre técnicas multivariadas en geografía física, incluye en el capítulo de clasificación un buen número de técnicas de ordenación (análisis de coordenadas principales, escalas multidimensionales y "Non-linear mapping"). Todas estas técnicas de ordenación tratan de representar geométricamente la estructura común o subyacente entre un conjunto de objetos, reduciendo el número de dimensiones exigidas para representar las relaciones entre los objetos y al mismo tiempo minimizar el grado de distorsión o "stress" dentro de estas relaciones. Todas las técnicas de ordenación tienen además la ventaja de poder aplicarse a datos categóricos (no-cuantitativos).

Otra técnica multivariada de creciente empleo es el análisis discriminante, que tiene un gran interés, ya que mientras el análisis de agrupamiento intenta descubrir los conglomerados "naturales" o agrupamiento en un conjunto de datos, el análisis discriminante asigna casos no adscritos a categorías a clases existentes. El uso conjunto del análisis de agrupamiento y discriminante constituye una combinación muy potente con aplicaciones de gran interés.

Además de las técnicas indicadas existen otras muchas que ofrecen un gran interés, pero que no han sido objeto de atención por los geógrafos. Así por ejemplo la técnica "*automatic interaction detection*" es un método de análisis de datos similar al análisis de agrupamiento, pero suministra más información sobre los lazos entre variables y la existencia de relaciones causales posibles. En este método aparece una variable dependiente y una serie de variables independientes (predictoras), medidas en cualquier tipo de escala aunque resulta más fácil de aplicar si la variable dependiente se mide en escala de intervalos. La técnica opera haciendo series de divisiones en los datos y agrupandolos en dos subconjuntos de acuerdo con un criterio basado en una de las variables independientes. La variable independiente que se emplea en cada división del

conjunto de datos es la que reduce al máximo el error de la suma de los cuadrados de los valores de los datos que se analizan. Es habitual presentar los resultados del método "automatic interaction detection" mediante un dendrograma que permite obtener tres tipos de información. Primero, una serie de subconjuntos que son de homogeneidad creciente. En segundo lugar, es posible detectar *niveles* de interacción entre las variables. Por último, puede captarse *la presencia* de interacción entre las variables.

En general, el método de "automatic interaction detection" es particularmente útil para descubrir las relaciones entre los datos y merece que los geógrafos cuantitativos le dediquen más atención en el futuro. Una buena exposición sobre las ventajas y posibilidades de este método se encuentra en el trabajo de Fielding, A. (1977).

Otra técnica poco utilizada por los geógrafos es el uso de *modelos de estructura latente*. Al igual que el análisis factorial, los modelos de estructura latente intentan clasificar fenómenos no observables tales como las nociones de *desarrollo*, *prejuicio racial*, *inteligencia*, *actitud con respecto al medio natural*, etc. Estos fenómenos que no pueden observarse directamente se llaman *variables latentes*, y se relacionan estrechamente con una o varias variables manifiestas. Sin embargo el análisis de estructura latente difiere del análisis factorial en varios aspectos importantes:

- a) el análisis de estructura latente no supone necesariamente relaciones lineales entre las *variables manifiestas y latentes*;
- b) el análisis de estructura latente no tiene el problema de determinar *comunalidades* como el análisis factorial; y
- c) el análisis de estructura latente es un conjunto de modelos, algunos de los cuales pueden utilizarse con datos de escala diferente a la de intervalos.

✓ Una revisión brillante y profunda sobre esta técnica pueden verse en Lazarfield, P.F. y Henry, N.W. (1968).

### El modelo lineal general

El decreciente interés por las técnicas multivariadas tiene una excepción importante en el modelo lineal general, especialmente los de regresión, correlación y análisis de varianza. A medida que los geógrafos se preocuparon por los trabajos de los estadísticos y econométricos, repararon en las enormes posibilidades de estas técnicas y de su aplicación. Los geógrafos han dedicado especial atención a subrayar la importancia que tiene el satisfacer los supuestos del modelo. Entre ellos cabe destacar a Poole, M.A. y O' Farrell, P.N. (1971). Otros autores como Mather, P. M. (1976), analizan variantes del modelo de regresión (ridge regression), que proporciona la solución parcial al problema de la *multicolinealidad*. Un segundo aspecto notable en el uso actual del modelo lineal general, especialmente el modelo de regresión lineal, es el empleo de programas interactivos de ordenador para obtener una formulación óptima del modelo en una situación de investigación específica. La idea no es nueva y sus antecedentes pueden rastrearse en la técnica de *regresión por etapas* utilizada en la selección de la combinación óptima de variables, a partir de un conjunto mucho mayor de variables, para incluirlas luego en una aplicación específica del análisis de regresión. En el análisis de regresión por etapas, una variable dada se incluye o desecha del grupo de variables independientes o explicativas en el modelo de regresión según un número de criterios muy limitados; en general, por el cambio en la suma de cuadrados explicada o por el test "t" o el test parcial F asociado a una variable; véase Draper, N.R. y Smith, H. (1966) si se desea ampliar detalles. Sin embargo, hay otros muchos criterios que pueden utilizarse en la selección y por ello no hay garantía de que se haya realizado una formulación óptima del modelo de regresión.

En los últimos años se han elaborado varias bibliotecas de programas tales como el MINITAB y el GLIM, que permiten construir el modelo de regresión específico de un modo interactivo, mediante el empleo del ordenador y teniendo en cuenta una amplia gama de criterios, incluyendo el conocimiento que tenga el investigador sobre el tema objeto de estudio. Estos procedimientos interactivos están muy influidos —como veremos más adelante— por las técnicas de exploración de análisis de datos. Un programa de

ordenador típico proporciona salidas impresas en pantalla, una información muy variada y diagramas que permiten al investigador estudiar el comportamiento de los datos y hacer decisiones subjetivas sobre su importancia. Los diagramas disponibles incluyen histogramas, diagramas de dispersión, representación de los residuales, etc. Esta información junto con las medidas utilizadas tradicionalmente (test "t" y F), proporcionan un método potente y flexible para determinar la forma del modelo de regresión. La idea fundamental de este enfoque de análisis de datos es como indican Henderson y Velleman, ¡"el investigador conoce mejor los datos que el ordenador"! HENDERSON, H.V. y VELLEMAN, P. F. (1981) p. 391. En otras palabras, en lugar de utilizar la pretendida *objetividad* del ordenador para decidir los rasgos esenciales del modelo, el analista de los datos toma una serie de decisiones *subjetivas* apoyadas en su conocimiento sobre el tema que investiga. En la práctica quiere decir que en lugar de uno o como mucho dos criterios, el investigador puede incluir muchos más criterios apoyándose en el comportamiento observado de los datos ó en su propia experiencia práctica y teórica.

Los modelos interactivos de regresión son un ejemplo interesante de convergencia de logros tecnológicos en la programación y en el análisis exploratorio de datos.

### Análisis de datos categóricos

En las secciones anteriores solo se presentaron métodos de análisis para emplear datos numéricos (medidos con escalas de intervalos y razones) aunque en algunos casos las técnicas se adaptaron a datos categóricos. En este apartado y en el siguiente nos centraremos en las técnicas diseñadas especialmente para analizar datos categóricos. (Vid. ESTEBANEZ, J.G. y BRADSHAW R.P.. (1979)). Conviene, no obstante, recordar que la división de los datos en dos tipos, así como la consiguiente diferenciación de dos tipos de técnicas estadísticas (paramétricas y no paramétricas), no siempre resulta tan clara como figura en los libros de texto de estadística. Existe una gran polémica sobre si las técnicas de correlación ordinal pueden utilizarse con datos ordinales, puesto que estas técnicas suponen que hay un número igual de intervalos entre los órdenes o rangos sucesivos, y por lo tanto, es preciso me-

## CUADRO I

### RELACION ENTRE TIPOS DE VARIABLES Y METODOS DE ANALISIS

	VARIABLES EXPLICATIVAS O INDEPENDIENTES			
VARIABLES RESPUESTAS	Continuas	Continuas (a)	Mixtas (b)	Categóricas (c)
O DEPENDIENTES	Categóricas	(d)	(e)	(f)

dir los datos mediante la escala de intervalos. Este punto de vista puede calificarse de purista. El punto de vista opuesto, el pragmático, argumenta que frecuentemente es mejor utilizar estadísticas paramétricas, incluso con datos que pueden considerarse más como de tipo ordinal que medidos con escalas de intervalos. Este punto de vista pragmático se apoya en que las técnicas paramétricas son más flexibles, más potentes e incluso más sensibles y tienen además un error de muestreo que es conocido (Vid. LABOVITZ, S. (1970)), y por lo tanto estas ventajas compensan cualquier error o violación de supuestos que puede surgir con el empleo de datos ordinales. Por lo tanto, como señala Doering y Hubbard a propósito de las siete categorías de la escala semántica diferencial, (siendo 1 muy positivo; 2— moderadamente positivo; 3— ligeramente positiva; 4— ni positivo, ni negativo, 5— ligeramente negativo; 6— moderadamente negativo; 7— muy negativo), que el purista argumenta que lo más adecuado es utilizar, en este caso, la escala ordinal, puesto que no hay confirmación empírica de que las siete categorías son equidistantes y por lo tanto, es preciso efectuar un análisis con métodos no —paramétricos (Doering, T.R. y Hubbard, R. (1979)). En el otro extremo, los paramétricos pueden aplicarse a los datos medidos por escalas ordinales, ya que es perfectamente razonable considerar estos datos como continuos y equidistantes. Así pues, el error en el que se incurra por aplicar tales métodos será despreciable, sobre todo al compararlo con las ventajas que se obtienen de su empleo.

De todo ello, la lección que los geógrafos han de recoger ante la dicotomía de lo paramétrico y no—paramétrico es adoptar una postura cualquiera entre el purismo extremo y el pragmatismo a ultranza, según la naturaleza del problema.

Por último, conviene advertir que se da con frecuencia una cierta confusión terminológica referida a las escalas de medida. Así por ejemplo, Wrigley emplea el término dicotómico en lugar de escala binaria; y policotomía desordenada para la escala nominal; y policotomía ordenada para la escala ordinal. Estas tres escalas se denominan de varias formas: categórica, no—paramétrica o no—métrica; asimismo también las escalas de intervalos y razones se conocen por los términos de escalas continuas, paramétricas o métricas.

A pesar de toda la problemática suscitada por las escalas y su definición en la década de los 70 se desarrolló una gran variedad de métodos de análisis de datos categóricos que pueden parangonarse ya con la producción ligada al modelo lineal general anteriormente tratado. Wrigley propone la clasificación que se incluye en Cuadro I de técnicas sobre el modelo lineal, (en este caso regresión) que pueden emplearse con diferentes tipos de datos (Ver Cuadro I).

En el caso de datos de tipo (a) son apropiados los métodos de regresión usuales como indican, entre otros Mather, P.P. (1976), o Johnston, R.J. (1978) o en los manuales de econometría como Kmenta, J. (1971). Para datos de tipo (b) y (c) también pueden utilizarse los métodos usuales con variables “dummy” (binarias).

Con datos de tipo (d), (e) y (f) la variable dependiente es categórica y requiere otros métodos analíticos. Wrigley, N. (1976) señala dos problemas cuando se aplica el modelo de regresión clásico a variables dependientes medidas con datos categóricos. En primer lugar, se viola el supuesto de homoscedasticidad que no resulta con ningún sesgo en la estimación de valores paramétricos, pero produce una pérdida de eficacia suscitando problemas serios cuando se emplean los test de inferencia convencionales. En segundo lugar, el modelo de regresión clásico de un valor estimado que puede variar entre  $-\infty$  y  $+\infty$ . Sin embargo, como el propio Wrigley indica, los valores estimados dentro del modelo categórico se interpretan mejor conforme la probabilidad oscila entre 0 y 1. La solución de este problema es sustituir el modelo de regresión convencional por un modelo alternativo basado en la función logística. En el caso de una variable dependiente de carácter dicotómico con variables independientes continuas (caso (d)), este modelo

$$\text{adopta la forma de } P_i = \frac{e^{f(x_i)}}{1 + e^{f(x_i)}} \quad \dots\dots\dots (1)$$

$$\text{siendo } f(x_i) = \alpha + \sum \beta_R X_{iR} \quad \dots\dots\dots (2)$$

y donde  $P_i$  representa la probabilidad de que se seleccione la primera categoría por el caso  $i$ , dados los valores de las variables explicativas  $K$ .

Como el valor de  $f(x_i)$  oscila entre  $-\infty$  a  $+\infty$  así también el valor de  $P_i$  oscila de 0 a 1.

El modelo logístico (1) puede expresarse de forma lineal así:  $\log_e \frac{P_i}{1-P_i} = f(x_i)$  ..... (3)

El primer término de la ecuación (3) es la transformación de  $P_i$  conocida con el nombre de transformación logit. La ecuación (3) es el modelo lineal logit y la ecuación (1) el modelo no-lineal logit. Wrigley, N. (1979) demuestra que las ecuaciones (1) y (3) puede aplicarse a variables dependientes policotómicas (nominales) y con mayor dificultad a variables explicativas mixtas (caso (e)) y a variables explicativas categóricas (caso (f)). A pesar de las dificultades apuntadas, el análisis es posible en todas estas circunstancias diversas (casos (d), (e) y (f)). Además, en los últimos años se ha producido un gran número de tests estadísticos semejantes al tipo "t" y un test de ajuste similar al  $R^2$ , del modelo convencional de regresión. Aunque se requiere una mayor investigación, sin embargo puede decirse que los resultados en este campo suponen una valiosa guía y ofrecen grandes posibilidades de aplicación.

Existen otros tipos de transformaciones que pueden igualmente emplearse, aunque por lo general, son difíciles de aplicar, y salvo el *modelo probit* pocas veces han sido utilizados en la práctica.

Afortunadamente existe un buen número de programas de ordenador que contienen procedimientos de análisis de datos categóricos como el XLOGIT y el ULOGIT adaptados a las casas CDC e IBM respectivamente, y el GLIM que es un programa del modelo lineal general que se encuentra en todas las Universidades británicas.

A pesar del carácter reciente de estas investigaciones en el campo de la Geografía, son ya relativamente numerosos los estudios en los que se aplica el análisis de datos categóricos a los problemas geográficos. Así por ejemplo, Wrigley desarrolló un modelo de representación cartográfica de una superficie de probabilidad que es el modelo logit equivalente a la representación cartográfica del análisis de superficies de tendencia

aplicado a los datos de tipo (d), (Cuadro I). Un buen estudio sobre las aplicaciones geográficas de esta técnica aparece en Wrigley, N. (1979).

### Medidas de Asociación

En esta sección consideraremos la aplicación de métodos análogos a los del modelo de correlación. Para uso con datos categóricos.

Las *medidas de asociación* son especialmente útiles en el análisis de tablas de contingencia. En la investigación geográfica se invierte mucho tiempo y esfuerzo en recoger y tabular datos, pero muy poco en analizar la información recopilada. Podemos utilizar las medidas de asociación para establecer la fuerza y dirección de la relación entre dos variables que aparecen en una tabla de contingencia. Por ejemplo podemos interesarnos por considerar la relación entre el color del cabello de las jóvenes y el color de sus ojos. Las medidas de asociación permiten determinar la fuerza de la relación, así como la dirección de la misma. Y así podemos interesarnos en conocer si la relación es en una sola dirección; por ejemplo si todas las jóvenes de ojos azules tienen el cabello rubio, pero que no todas las rubias tienen ojos azules. O bien en dos direcciones; que todas las jóvenes de ojos azules tienen el pelo rubio y que todas las rubias tiene los ojos azules.

Las medidas de asociación son un poderoso y útil medio de análisis de la información presente en las tablas de contingencia. Desgraciadamente, la situación es muy complicada por la proliferación de medidas de asociación disponibles. Esta circunstancia se debe, en parte, a que muchas medidas se idearon para circunstancias diferentes. De esta forma, medidas como la T de Tschuprow, V de Cramer o el coeficiente de contingencia C de Pearson, fueron ideados para tratar datos nominales, mientras que los coeficientes  $\tau_{ab}$  y  $\tau_c$  o el D de Somers lo fueron para datos ordinales. De igual modo, el coeficiente Q de Yule, o el  $f_1$  lo son para aplicar a tablas de contingencia de  $2 \times 2$ , en tanto que los coeficientes tau de Goodman y Kruskal y la medida de asociación lambda se idearon para el tratamiento de cualquier tipo de tablas de contingencia. (Véase MEULER, J.H. y SCHUESSLER, K.F. (1961) y BLALOCK, H. (1961)) para un tratamiento más detallado de estas medidas de asociación). La se-

gunda razón que explica la proliferación de medidas de asociación es que muchas de ellas se han concebido para extraer diferentes tipos de información de las tablas de contingencias.

Se han hecho varios intentos de producir una única medida de asociación. Por ejemplo Leik, R.K. y Gove, W. (1971) conciben un coeficiente  $d_n$  susceptible de aplicación a datos o medidas en escalas nominales, ordinales y de intervalos. Sin embargo, no hay certeza de que esta medida suponga un avance sobre las medidas existentes y por ello los investigadores aplican el coeficiente que consideran más idóneo al problema objeto de estudio.

Debe tenerse especial cuidado al usar medidas tales como la V de Craner, la C de Pearson y la T de Tschuprow, que se basan en el coeficiente de la Ji Cuadrado. Como ponen de manifiesto Goodman, L.A. y Kruskal, W.H. (1970) p. 740, "el hecho de que la ji cuadrado sea un test excelente de independencia no significa que la ji cuadrado, o una función de la ji cuadrado, haga automáticamente medidas de asociación a pesar de su gran difusión en los estudios empíricos".

#### Modelos causales: El método de Simon—Blalock y el Path Analysis

Sorprende el escaso interés que suscitó el problema de la causalidad en la estadística clásica. La mayoría de las técnicas convencionales suponen relaciones causales de un modo simple; y así en la correlación simple se supone que Y es causada por X y en la regresión múltiple Y es causada por  $X_1 + X_2 + X_3 + \dots + X_n$ .

Sin embargo, estas situaciones son muy simples y en la mayoría de los casos en geografía, y especialmente en geografía humana, hemos de tratar con problemas mucho más complejos. Los modelos causales intentan comprender estas situaciones complejas y, cuando se emplean correctamente, pueden ayudarnos a desarrollar modelos más efectivos y decidirse entre diferentes modelos competitivos, por el más adecuado y congruente con los datos empíricos.

Los modelos causales son técnicas más complejas que las de correlación o regresión sim-

ple, y además son menos precisas y no puede decirse que modelo específico proporcional el "mejor ajuste" a un conjunto de datos. La técnica de los modelos causales se resiente también, de limitaciones prácticas, en el sentido de que es difícil estudiar las relaciones entre más de seis variables por la lentitud y tediosidad del cálculo.

Llegamos a este punto, tal vez sea útil aclarar el significado de *causalidad*. La causalidad se emplea normalmente al referirnos a un proceso que conduce inevitable y lógicamente a un resultado específico. De esta forma, cuando decimos que Y es causada por X, queremos decir que al aparecer X se sigue una serie de sucesos que inevitablemente conducen a Y. Sin embargo, en la estadística no es posible decir que X causa Y, sino que la evidencia empírica es consistente con la noción de que X causa Y. Disponemos de un número de principios que nos ayudan a determinar los hechos que pueden estar ligados mediante un proceso causal. Estos principios son la *aparición conjunta*, donde si aparece X también aparece Y; el principio de *covarianza* según el cual un cambio en X lleva a un cambio en Y; el *principio de producción*, donde hay una posibilidad práctica de que la causa que se supone bajo hipótesis produce el efecto conocido. Por ejemplo, la lluvia puede causar inundaciones, pero no causa el periodo seco que sigue a las inundaciones. En los modelos causales, se establecen las relaciones para uno de estos cuatro principios, el *principio de covarianza*. Por este motivo los modelos causales no han de considerarse como guía global de relaciones causales, sino más bien como una técnica exploratoria que suministra hipótesis de trabajo y nunca verdades constatadas.

Con objeto de especificar los modelos causales es normal incluir una serie de diagramas en los que se señala la dirección hipotética de la relación causal mediante una flecha (Fig. 1). En esta figura se presentan las cuatro relaciones posibles entre dos variables: a) X causa Y; b) Y causa X; c) X causa Y y al mismo tiempo Y causa X, es decir se trata de una relación de retroalimentación (feedback) y sólo se puede medir este tipo de relación usando series de datos temporales; d) X e Y no están causalmente relacionadas y por ello se omite la flecha. Cuando se trata de más de dos variables los modelos se hacen más complejos, como se indica en la Fig. 2. Una de las



Fig. 1

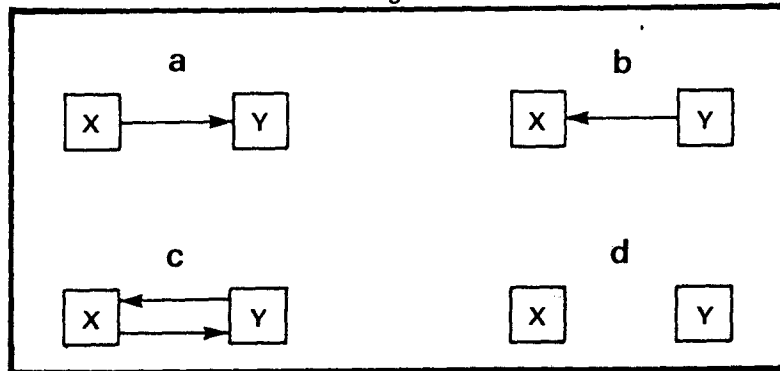


Fig. 2

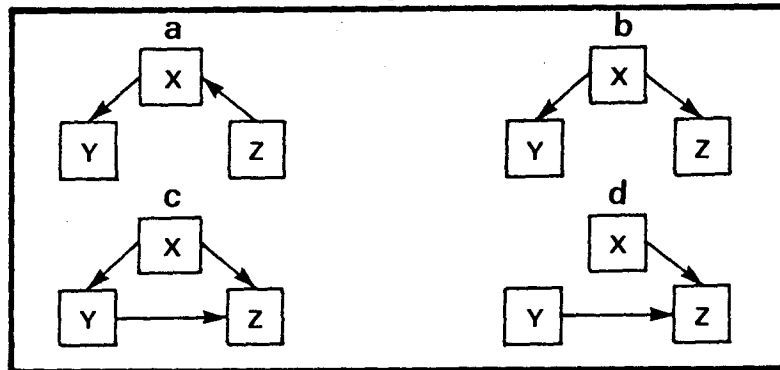


Fig. 3

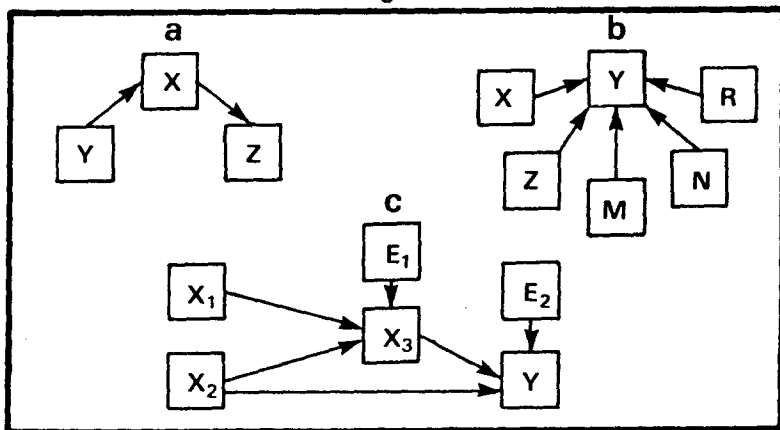


Fig. 1: Todas las posibles relaciones causales entre pares de variables

Fig. 2: Algunas de las posibles relaciones causales entre tres variables

Fig. 3: Comparación entre un modelo Simon-Ballock (a) y dos modelos del tipo "path analysis" (b) y (c).

ventajas de los modelos causales es que ayudan a establecer y diferenciar correlaciones reales de las espúreas. Así en el caso señalado en la Fig. 2 (b) se muestra una alta correlación aparente entre las variables Y y Z, en tanto que esta correlación está producida por una tercera variable (X).

Como se ha apuntado anteriormente no existe garantía de que ningún modelo proporcione el "mejor ajuste" a un conjunto de datos empíricos. Una posible solución es especificar todos los modelos posibles y ver cuál de ellos suministra el "mejor ajuste", pero desgraciadamente esta solución sólo es factible cuando se analiza un número muy pequeño de variables (2 ó 3). Como subraya Pringle, un pequeño aumento en el número de variables origina un incremento considerable en el número de modelos, así que con 6 variables el número total de modelos diferentes puede exceder la cifra de varios miles (PRINGLE, D.G. 1980, p. 7). Por esta razón los modelos causales no han de multiplicarse indebidamente y deben ser lo más simples posibles. Siempre que sea posible disponer de otra información debería utilizarse para establecer la forma y dirección de las relaciones causales, permitiendo de este modo que el análisis causal se centre en las partes desconocidas de problema de investigación.

El análisis causal de sendas, o *path analysis* es semejante al sistema de modelos propuestos por Simon y Blalock, en el sentido de que ambas técnicas tratan de comprender las relaciones causales presentes en el modelo. Sin embargo existen algunas diferencias notables. La técnica propuesta por Simon y Blalock exige coeficientes de correlación parcial, en tanto que el "path analysis" utiliza coeficientes estandarizados de regresión parcial. En segundo lugar, el enfoque de Simon-Blalock examina toda una serie de causas efectos, en tanto que el "path analysis" examina el efecto de un número de variables sobre una variable dependiente. La figura 3 (a) y la Figura 3 (b) comparan los modelos de Simon-Blalock y el "path analysis" respectivamente. En tercer lugar, el "path analysis" supone un modelo y por lo tanto estima los valores de los parámetros, en tanto que el enfoque de Simon-Blalock comprueba si el modelo hipotético es consistente con los datos empíricos. En cuarto lugar, en el "path analysis" las *variables exógenas* (variables

que se especifican pero cuyas causas se excluyen del modelo) y *variables endógenas* (variables cuyas causas se incluyen dentro del modelo), pueden analizarse en el mismo modelo. Así pues en la Fig. 3 (c)  $X_1$  y  $X_2$  influyen en  $X_3$ , y  $X_2$  y  $X_3$  influyen en Y. En este diagrama  $X_1$  y  $X_2$  son *variables exógenas* y  $X_3$  e Y son *variables endógenas*.  $E_1$  y  $E_2$  son *variables residuales* que no se especifican en el modelo pero en el caso de que se incluyan explicarán la varianza que queda en las *variables endógenas*.

Hasta aquí hemos hecho el supuesto implícito de que los modelos causales exigían datos medidos en escalas de intervalos (para obtener los coeficientes de correlación y regresión). Sin embargo puede generalizarse el método al análisis de datos categóricos. Por ejemplo, Davis proporciona un estudio global de como emplear el coeficiente Q de Yule en el análisis de las relaciones causales entre variables binarias (DAVIS, J.A. (1971).

El modelo causal representa un conjunto poderosos de técnicas que puede emplearse para descubrir muchas relaciones complejas que se estudian en geografía. El método de Simon-Blalock y el "path analysis" forman parte del modelo lineal general y ambos están sometidos a los supuestos del modelo, así como a la limitación del número de variables que pueden analizarse (PRINGLE, D.G. (1980) pp. 30-33). A pesar de las limitaciones de los modelos causales, éstos prometen popularizarse en el futuro inmediato en la investigación geográfica y de un modo especial en la exploración de datos.

### Análisis espacial y series temporales.

Al comenzar la década de los años 70, la mayoría de los geógrafos consideraban el análisis de series de datos espacio-temporales como algo tangencial en la geografía cuantitativa. La autocorrelación espacial se consideró como un problema que había de tenerse en cuenta a la hora de aplicar el análisis de regresión a los datos espaciales. El análisis de series temporales solo fue considerado interesante por los econométricos. Sin embargo, pronto se reparó que la autocorrelación espacial lejos de ser un problema que exigía un tratamiento, ha de ser el objeto central de investigación en muchos estudios geográficos. Es decir, en muchos casos nos interesamos espe-

cificamente en saber si la distribución espacial de un fenómeno es aleatoria o no. Gran parte del trabajo pionero en esta línea de investigación se debe a Ord y Cliff (CLIFF, A.D. y ORD, J.K. 1980).

En los últimos años aparece un interés creciente por los cambios ocurridos en el tiempo, especialmente en geografía económica, física e histórica. La mayor parte del interés se centra en torno a la aplicación de los modelos de Box—Jenkins. En la pasada década gran parte de la atención se centró en problemas técnicos, sin embargo en el momento presente ésta se dirige hacia la aplicación de estas técnicas (HEPPLE, L.W. (1981) pp. 94—95).

Es interesante subrayar que en la actualidad se hacen serios intentos para aunar estas dos líneas de investigación en un cuerpo integrado de "análisis espacio—temporal". Por ejemplo, Cliff y Ord proponen el empleo de un índice espacio—temporal para comprobar la autocorrelación espacio—temporal (CLIFF, A.D. y ORD, J.K. (1981). Este esfuerzo parece complementarse muy bien con las importantes aportaciones hechas por Hägerstrand y colaboradores en la geografía del tiempo. El futuro de esta nueva línea promete marcar una auténtica revolución en la geografía tradicional y de un modo especial en la humana y regional (CARLSTEIN, T., PARKES, D.N. y THRIFT, N.J. (1977)).

#### Estadística Bayesiana.

Durante mucho tiempo este campo de la estadística mereció escasa atención por parte de geógrafos, sin embargo es muy probable que en un futuro inmediato se corrija esta tendencia, porque el enfoque bayesiano ofrece un campo de aplicación de gran interés en toda la geografía y especialmente en la geografía del comportamiento.

Antes de proseguir, conviene señalar, a pesar de la tendencia que intenta enfrentar el enfoque bayesiano y el clásico y presentarlos como irreconciliables, que el análisis estadístico bayesiano intenta alcanzar objetivos diferentes a los que plantea el análisis estadístico clásico, por lo que ha de verse como un enfoque complementario y no como un sistema rival y opuesto.

La estadística bayesiana aparece con un ar-

tículo del Reverendo Tomás Bayes, publicado tras su muerte por un grupo de amigos (BAYES, T. (1763)). En este artículo Bayes intenta invertir la relación probabilística, aceptada normalmente, entre la hipótesis y el resultado. En las pruebas de hipótesis clásicas, se formula una hipótesis y se calcula la probabilidad de ciertos resultados. En el análisis bayesiano se parte de un resultado determinado y se calculan las probabilidades de varias hipótesis.

Durante los siglos XVIII y XIX, matemáticos y estadísticos como Laplace, Gauss, Galton y Edgeworth prestaron un gran interés al teorema bayesiano, aunque continuaron trabajando predominantemente siguiendo la tradición clásica. Sin embargo, a comienzos del S. XX, Sir Ronald Fisher rechazó explícitamente el enfoque de Bayes y trató de sustituirlo por el sistema de comprobación de la hipótesis dentro de la tradición clásica. El trabajo de Fisher condujo al procedimiento de Neyman—Pearson de comprobar la hipótesis de investigación oponiéndola a la hipótesis nula, con la aceptación consiguiente de cometer un error de Tipo I, (rechazando la prueba cuando es verdadera), o de Tipo II (no rechazando la prueba cuando es falsa).

La oposición de Fisher y el éxito generalizado del procedimiento de Neyman—Pearson supuso un cierto olvido por parte de los estadísticos al análisis de Bayes, hasta la década de los años 60. El cambio de actitud se originó por un cierto desencanto con los procedimientos clásicos de comprobación de hipótesis, así como por los trabajos de Jeffreys y Savage sobre la teoría estadística de decisiones que permitió introducir los conceptos de probabilidad y beneficios subjetivos en el análisis bayesiano. El procedimiento de la comprobación de la hipótesis de Neyman—Pearson suscita un buen número de dificultades prácticas siendo la más importante la referente al valor del parámetro que ha de comprobarse en la hipótesis nula. Por ejemplo, en el análisis de correlación es usual ajustar el parámetro de la población a cero y comprobar si el resultado muestral se encuentra próximo a este valor. Si se rechaza la hipótesis nula, eliminamos entonces la posibilidad de que el parámetro poblacional tenga un valor de cero, pero deja todavía un número infinito de posibilidades de obtener otros valores. Por otra parte, de no rechazar la hipótesis nula,

entonces el valor de la población cero es uno de los muchos valores posibles del parámetro.

La obra de otros estadísticos desde 1940, y de un modo especial de L. Savage sobre la teoría de decisiones, permitió que todas las cuestiones sobre probabilidad personal se analicen mediante las técnicas propuestas por Bayes (SAVAGE, L. J. (1972)). En esta formulación la cuestión estadística es que dada una situación, cómo debería comportarse una persona, teniendo en cuenta que tal persona sigue siendo fiel a sus creencias y preferencias personales. De esta forma, el análisis bayesiano se generalizó e incluye las pruebas estadísticas de teorías normativas de comportamiento humano.

Hasta el momento presente la mayoría de los estudios dentro de la geografía del comportamiento se llevaron a cabo dentro de la estructura descriptiva-conductista y la mayoría de los investigadores, siguiendo a Simon, rechazan el concepto de "hombre económico optimizante" sustituyéndolo por el modelo de hombre con fines "satisfactorios". No obstante, este último modelo de hombre con fines satisfactorios pocas veces ha sido comprobado en la práctica y en la mayoría de los casos sólo se puede establecer el modelo citado mediante la observación de los resultados de conductas pasadas. Con el enfoque bayesiano normativo puede centrarse la atención en el problema de tomar decisiones racionales ante situaciones de incertidumbre. Este enfoque proporciona a la geografía del comportamiento un conjunto de herramientas analíticas más poderosas para elaborar modelos teóricos y resolver problemas prácticos.

El análisis bayesiano está constituido por tres distribuciones de probabilidad: la *distribución anterior* que puede estar formada por una visión o preferencia personal; la *distribución condicional* o de predicción, que está formada de evidencia empírica, datos muestrales, etc., referentes al tema de estudio; y la *distribución posterior* que define el estado más probable del problema objeto de estudio, según la información contenida en las *distribuciones anterior, y condicional*. Expresado de manera informal, la relación sería:

Distribución posterior  $\propto$  (Distribución anterior)

(Distribución condicional).

Siendo  $\alpha$  proporcional a/o igual a, siendo a un valor constante.

El cálculo de la estadística bayesiana puede ser muy complicado, aunque existe una familia de distribuciones beta que sirven con bastante aproximación para calcular cualquier *distribución anterior o condicional* que puedan aparecer en la práctica. Otro punto a subrayar en el cálculo bayesiano es que en determinadas circunstancias es posible usar *distribuciones posteriores* como una estimación mejorada sobre el estado del problema objeto de estudio, que puede emplearse a su vez como una versión mejorada de la *distribución anterior* y utilizarse luego en la última fase del estudio.

En conclusión, puede decirse que el análisis bayesiano proporciona un sistema de análisis poderoso y flexible con muchas aplicaciones en geografía. Una revisión sobre este enfoque puede verse en las obras de PHILIPS, L.D. (1973) y BOX, G.E.P. (1973), así como en la obra de HAYTER, R. (1975) en la que presenta interesantes aplicaciones en la geografía del comportamiento.

#### Métodos más generales de análisis cuantitativo

Todas las técnicas presentadas hasta este momento se apoyan en principios estadísticos. Ahora bien, la geografía cuantitativa cubre un campo mucho más amplio. Por ejemplo, durante la última década, se realizó un gran esfuerzo en el desarrollo operativo de modelos urbanos y regionales. La mayoría de estos modelos se inspiran en los modelos ideados en los años 60 por Lowry, si bien A.G. Wilson y sus colaboradores también intentaron rechazar la desacreditada "Física social" apoyándose en la formulación de sus modelos en el concepto de entropía (WILSON, A.G. (1974)). Asimismo se hicieron importantes logros en el análisis de puntos, en la teoría de grafos, etc. (RIPLEY, B.D. (1981)). A estos logros es preciso añadir técnicas nuevas y revolucionarias para analizar problemas geográficos, especialmente la *teoría de las catástrofes* y el *análisis Q*.

Sin que podamos formular conclusiones definitivas, cabe señalar que se abre una era en la que las técnicas se alejan de los supuestos del

**CUADRO II**

**MATRIZ DE CORRELACION – EJEMPLO DE UNA SALIDA DEL ORDENADOR**

**MATRIZ DE CORRELACION**

	1	2	3	4	5	6	7	8
1	1.0000							
2	0.0270	1.0000						
3	0.2510	0.3792	1.0000					
4	0.5466	0.4716	0.3377	1.0000				
5	-0.4485	-0.5136	-0.8923	-0.4880	1.0000			
6	0.2776	0.5049	0.5917	0.3486	-0.7959	1.0000		
7	-0.2846	-0.2510	-0.7554	-0.3307	0.7438	-0.3958	1.0000	
8	0.3434	0.6605	0.8658	0.5249	-0.9088	0.6194	-0.7522	1.0000
9	-0.5399	-0.5073	-0.8018	-0.4997	0.8126	-0.4428	0.7670	-0.8223
10	-0.1429	-0.2186	-0.7380	-0.0466	0.7532	-0.6007	0.6291	-0.6013
11	0.2505	0.4789	0.8291	0.4129	-0.8940	0.7037	-0.6867	0.7976
12	0.3022	0.4801	0.8526	0.4015	-0.8877	0.7792	-0.7323	0.7987
13	-0.5524	-0.3175	-0.3040	-0.6237	0.3643	-0.2687	0.2024	0.4373
14	-0.2140	-0.4821	-0.5475	-0.3447	0.4966	-0.3801	0.1509	-0.6231

**MATRIZ DE CORRELACION**

	9	10	11	12	13	14
9	1.0000					
10	0.6133	1.0000				
11	-0.7203	-0.7353	1.0000			
12	-0.7491	-0.7154	0.8540	1.0000		
13	0.4374	0.0028	-0.2164	-0.4067	1.0000	
14	0.3438	0.1975	-0.3641	-0.4455	0.6476	1.0000

### CUADRO III

#### NOMBRE DE LAS VARIABLES

1. Incremento de la población 1960-70.
2. Densidad de población 1970 (hab/km<sup>2</sup>).
3. Nivel de urbanización (porcentaje de la población en localidades de 10.000 y más).
4. Población activa (de 12 años y más).
5. Población activa en agricultura.
6. Población activa en industria.
7. Porcentaje de analfabetos (mayores de 10 años).
8. Porcentaje por ocupación principal: profesionales, técnicos y funcionarios.
9. Porcentaje de ingresos mensuales con menos de 500.000 pesos.
10. Porcentaje de viviendas con un solo cuarto.
11. Porcentaje de viviendas que disponen de agua corriente.
12. Porcentaje de ocupantes de viviendas que tienen energía eléctrica.
13. Porcentaje de viviendas donde los ocupantes no han comido carne en la semana anterior al censo.
14. Promedio de hijos por mujer.

FUENTE: CENSO 1970 de los Estados Unidos de México.

Fig. 4

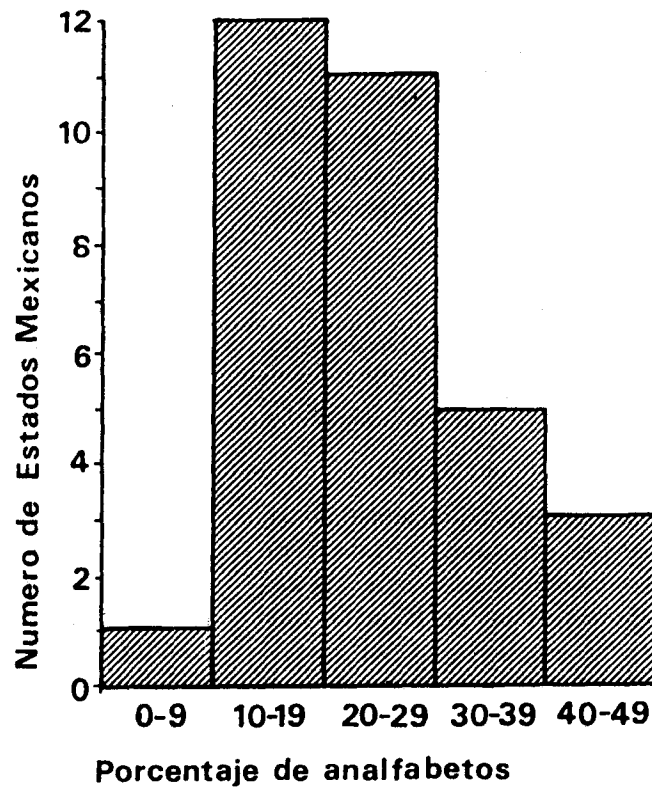


Fig. 4: Histograma de los analfabetos en los Estados Mexicanos.

modelo lineal general y se dirigen a un camino en el que se "deja que los datos hablen por ellos mismos". Una revisión de estas técnicas puede verse en los trabajos recientes de Wilson, A.G. (1981) y Beaumont, J.R. y Gatrell, A.G. (1982). También se dió gran importancia a problemas tales como los de aplicar métodos cuantitativos en geografía en relación a unidades espaciales modificables y al manejo de grandes masas de datos. El primer problema ya tiene una larga historia y se conoce con el nombre de falacia ecológica. Es Openshaw el que más contribuyó al estudio y resolución de este problema aunque sin resultados plenamente satisfactorios (OPENSHAW, S. y TAYLOR, P.J. (1981)).

En lo que se refiere al manejo de grandes conjuntos de datos, puede consultarse la obra de Devereux, B.J. (1978), pues aunque no afecta los métodos cuantitativos tradicionales, forma el nudo gordiano de campos afines a la geografía cuantitativa como el de la cartografía automática y el análisis de los datos de las imágenes proporcionadas por sensores remotos.

#### Análisis exploratorio de datos

De lo expuesto en las anteriores secciones, se deduce que el análisis exploratorio de datos es una de las principales tendencias en la geografía cuantitativa.

Los métodos de análisis exploratorio de datos proporcionan dos aspectos muy positivos. En primer lugar subrayan todo lo que atañe a la mejor comprensión de los datos, en lugar de quedarse solamente en la comprensión del modelo como ocurrió en la geografía cuantitativa en la década de los 60. En segundo lugar, este análisis de datos ayuda a seleccionar qué modelo o técnica multivariada es la adecuada y la más idónea para tratar un conjunto de datos. En este campo prometedor, es preciso destacar la obra de John Tukey, ya que este autor ideó un gran número de técnicas de análisis exploratorio de datos (TUKEY, J. (1977)). Nosotros presentaremos aquí dos técnicas que consideramos de gran utilidad.

Como es bien sabido, el histograma es la forma más sencilla de presentar un conjunto de datos. La Fig. 4. es un histograma que representa la distribución de analfabetos en los 32 estados de México. En esta distribución no se indica el

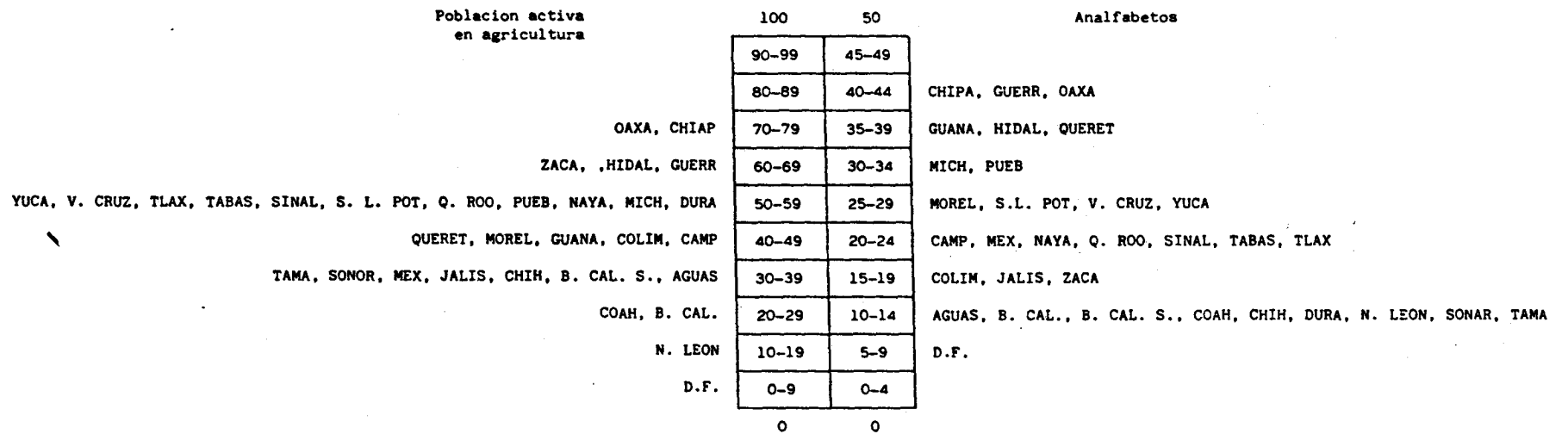
comportamiento individual de cada estado. Tukey a fin de superar esta limitación desarrolló una serie de procedimientos de presentación de datos. En la Fig. 5 se muestran dos diagramas. En el diagrama situado a la derecha se representan los analfabetos, es idéntico al diagrama de la Fig. 4, salvo en que en éste último se identifica cada estado, lo que facilita la evaluación porcentual de cada estado. Al comparar dos variables se puede observar la posición relativa de los estados con respecto a las variables que se examinan. Así por ejemplo, en la Fig. 5, se aprecia que los estados de Chiapas, Guerrero y Oaxaca tienen un alto porcentaje de población activa agraria y un alto porcentaje de analfabetos, en tanto que el Distrito Federal y Nuevo Leon tienen los dos valores muy bajos en las dos variables. En contraste Zacatecas tiene un elevado porcentaje de población activa agraria pero un valor muy bajo de analfabetos.

Otra forma muy común y generalizada de presentar la relación entre variables es mediante una matriz de coeficientes de correlación. Sin embargo no resulta sencillo identificar todas las relaciones importantes en los datos. Por ejemplo, el cuadro II presenta la matriz de correlaciones entre 14 variables en 32 estados mexicanos (nombres de las variables están contenidos en cuadro III). De esta matriz se deduce fácilmente que la variable 8 se relaciona muy fuerte con la 3 y con la variable 5. Sin embargo la información así deducida es lenta y no hay garantía de recoger toda las relaciones posibles en los datos. Ahora bien, simplificando los valores y presentándolos en una matriz completa (Cuadro IV), resulta más fácil la interpretación de la matriz. Otra mejora se consigue reordenando las variables con objeto de buscar los agrupamientos naturales dentro de los datos. Esto puede hacerse cortando la matriz en filas separadas y colocando la fila con los valores negativos mas elevados en la parte inferior de la tabla. Las restantes filas se colocan luego entre estas dos extremas. El proceso se repite con las columnas hasta conseguir una matriz simétrica (Cuadro V). En la mayoría de los casos este procedimiento lleva a una matriz que es muy fácil de comprender e interpretar. Por ejemplo en el Cuadro V puede observarse que las variables 8,3,12,11,6,2,4 y 1 están todas ellas correlacionadas entre sí positiva y negativamente con las restantes variables.



Fig. 5: "Rama y hoja" diagrama de analfabetos y la población activa en agricultura en los Estados Mexicanos.

Fig. 5



CUADRO IV

## MATRIZ DE CORRELACION – SIMPLIFICADA

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1														
2	.03													
3	.25	.03												
4	.55	.47	.25											
5	-.45	-.51	-.89	-.49										
6	.28	.50	.59	.35	-.80									
7	-.28	-.25	-.76	-.33	.74	-.40								
8	.34	.66	.87	.52	-.91	.62	-.75							
9	-.54	-.31	-.80	-.50	.81	-.44	.77	-.82						
10	-.14	-.22	-.74	-.05	.75	-.60	.63	-.60	.61					
11	.25	.48	.83	.41	-.89	.70	-.69	.80	-.72	-.74				
12	.30	.48	.85	.40	-.89	.78	-.72	.80	-.75	-.72	.85			
13	-.55	-.32	-.30	-.62	.36	-.27	.20	-.44	.44	.00	-.22	-.41		
14	-.21	-.48	-.55	-.34	.50	-.38	.15	-.62	.34	.20	-.36	-.45	.65	

CUADRO V

MATRIZ DE CORRELACION – SIMPLIFICADA Y REORDENADA

	8	3	12	11	6	2	4	1	13	14	10	9	7	5
8		.87	.80	.80	.62	.66	.52	.34	.44	-.62	-.60	-.82	-.75	-.91
3	.87		.85	.83	.59	.38	.34	.25	-.30	-.55	-.74	-.80	-.76	-.89
12	.80	.85		.85	.78	.48	.40	.30	-.41	-.45	-.72	-.75	-.72	-.89
11	.80	.83	.85		.70	.48	.41	.25	-.22	-.36	-.74	-.72	-.69	-.89
6	.62	.59	.78	.70		.50	.35	.28	-.27	-.38	-.60	-.44	-.40	-.80
2	.66	.38	.48	.48	.50		.47	.03	-.32	-.48	-.22	-.31	-.25	-.51
4	.52	.34	.40	.41	.35	.47		.55	-.62	-.34	-.05	-.50	-.33	-.49
1	.34	.25	.30	.25	.28	.03	.55		-.55	-.21	-.14	-.54	-.28	-.45
13	-.44	-.30	-.41	-.22	-.27	-.32	-.62	-.55		.65	.00	.44	.20	.36
14	-.62	-.55	-.45	-.36	-.38	-.48	-.34	-.21	.65		.20	.34	.15	.50
10	-.60	-.74	-.72	-.74	-.60	-.22	-.05	-.14	.00	.20		.61	.63	.75
9	-.82	-.80	-.75	-.72	-.44	-.31	-.50	-.54	.44	.34	.61		.77	.81
7	-.75	-.76	-.72	-.69	-.40	-.25	-.33	-.28	.20	.15	.63	.77		.74
5	-.91	-.89	-.89	-.89	-.80	-.51	-.49	-.45	.36	.50	.75	.81	.74	

De un modo semejante, puede verse que las variables 8,3,12 y 11 están más altamente correlacionadas entre sí que las variables 6,2,4 y 1. En el otro extremo del Cuadro V, puede verse que las variables 10, 9,7 y 5 están altamente correlacionadas, en tanto que la 13 y 14 lo están mucho menos.

## CONCLUSIONES

Durante el último decenio el estudio de la geografía ha sufrido una transformación notable. Ha sido tal vez, la década más apasionante e innovadora de la historia geográfica. En lo que atañe a la geografía cuantitativa se ha producido un florecimiento extraordinario de nuevas ideas y métodos, como hemos analizado. Del examen realizado ¿qué puede decirse sobre el futuro de la geografía cuantitativa?.

En el panorama que hemos presentado cabe enumerar las tendencias que más han de influir en el desarrollo de la disciplina en la década futura. Estas son:

a) El número creciente y la complejidad de las técnicas cuantitativas hace que la *especialización* en el campo de la geografía cuantitativa sea inevitable. En el futuro inmediato ningún geógrafo cuantitativo será capaz de estar al día y de asimilar todos los avances que experimente la disciplina en este campo.

b) Debido a la especialización creciente será preciso dedicar más atención a la *organización* de los geógrafos cuantitativos. Existe el peligro de que el especialista, aislado en su universidad, pierda el contacto con los avances en su disciplina, a no ser que se reúnan e intercambien ideas e información periódicamente. Estos encuentros no deberían reducirse al ámbito español, sino también promover reuniones con los geógrafos de otros países, así como con otros colegas pertenecientes a otros campos como matemáticos, estadísticos, programadores, etc.

c) La próxima década se caracterizará más por la *consolidación y empleo* de las técnicas existentes que por el desarrollo de otras nuevas líneas y áreas de análisis.

d) Con el descenso continuo de los precios

reales de los ordenadores, los geógrafos podrán disponer de estos instrumentos con gran facilidad. Especialmente de los *micro-ordenadores* que son capaces hoy en día, de analizar el 90 % de los problemas que se suscitan en la geografía cuantitativa.

e) La tendencia en la industria de los ordenadores es a crear *sistemas de programas de fácil uso*. Por lo tanto, en el futuro se dispondrá de un número creciente de técnicas que puede utilizar el no especialista.

f) La tendencia, dentro de la programación es crear *programas interactivos* que permiten al investigador alimentar y dar respuestas y decisiones durante la ejecución del programa.

g) Se dará mas importancia a los *problemas que susciten los datos* que al empleo de técnicas complejas. Por lo tanto, el uso de análisis exploratorio de datos o la preocupación por los problemas que originan las unidades especiales modificables, serán líneas de investigación de gran importancia.

h) Existe una preocupación creciente por el uso de técnicas cuantitativas para *resolver problemas* más que estudiar las técnicas en sí mismas.

i) Hay mayor interés por los lazos entre las técnicas cuantitativas y los *logros teóricos* en el resto de la geografía. La geografía cuantitativa se considera cada vez más como un campo de investigación que esta contribuyendo en gran medida al desarrollo de otras ramas de la geografía y no como una disciplina separada que intenta crear su propia ciencia espacial.

Considero que todas estas tendencias son muy saludables para la geografía cuantitativa y que en los próximos años ofrecerán perspectivas muy sugerentes e innovadoras. En efecto, no carecemos de "problemas espaciales" para que el geógrafo cuantitativo investigue y trate de comprenderlos. Parece verosímil que el geógrafo cuantitativo en el próximo decenio desempeñará un creciente e importante papel en el desarrollo de la geografía tanto en sus aspectos teóricos como aplicados.

## BIBLIOGRAFIA

- ANDERBERG, M.R. (1973): "Cluster analysis for applications". New York, Harcourt, Brace, Jovanich.
- BAYES, T. (1763): "Essay towards solving a problem in the doctrine of chances". *Philosophical Transactions of the Royal Society (London)* Vol. 53 páginas 370-418.
- BEAUMONT, J.R. y GATRELL, A.C. (1982): "An introduction to Q-analysis". (CATMOG 34) Norwich, Geo Abstracts.
- BERRY, B.J.L. (1971): "Introduction: the logic and limitations of comparative factorial ecology" *Economic Geography* Vol. 47 num. 2 (supplement), p. 209-219.
- BOX, G.E.P. (1973): "Bayesian inference in statistical analysis". Reading, Mass., Addison-Wesley.
- CARLSTEIN, T. PARKES, D.N. y THRIFT, N. J. (1977): *Timing space and spacing time in socio-economic systems*. London, Arnold.
- CLIFF, A.D. y ORD, J.K. (1980): *Spatial processes: models and applications*. London, Pion.
- CLIFF, A.D. y ORD, J.K. (1981): "Spatial and temporal analysis: autocorrelation in space and time" p. 104-110 en N. WRIGLEY y R.J. BENNETT *Quantitative Geography*. London, Routledge Kegan and Paul.
- DAVIS, J.A. (1971): *Elementary survey analysis*. Englewood Cliffs, Prentice-Hall.
- DEVEREUX, B.J. (1978): "Grip—a geographical retrieval and information processing package". *Computer Application* vol. 5 num. 1 + 2 p. 839-859.
- DOERING, T.R. y HUBBARD, R. (1979): "Measurement and statistics: the ordinal interval controversy and geography". *Area* vol. 11 num. 3 p. 237-243.
- DRAPER, N.R. y SMITH, H. (1966): "Applied regression analysis". New York, Wiley.
- ELFERS, H. (1980): "On uninterpretability of factor analysis results" *Transactions, Institute of British Geographers (New Series)* Vol. 5 num. 3 p. 318-329.
- ESTEBANEZ, J.G. y BRADSHAW, R.P. (1979): *Técnicas de cuantificación en geografía* Madrid, Tebar Flores.
- FIELDING, A. (1977): "Binary segmentation: the automatic interaction detector and related techniques for exploring data structure" p. 221-258 en C.A. O' MUIRCHEARTAIGH y C. PAYNE *Exploring data structures*, volume 1, Chichester, Wiley.
- GIGGS, J.A. y MATHER, P.M. (1975): "Factorial ecology and factor invariance: an investigation". *Economic Geography* Vol. 51 p. 366-382.
- GOODMAN, L.A. y KRUSKAL, W.H. (1954): "Measures of association for cross classification" *Journal of the American Statistical Association* Vol. 49 p. 732-764.
- GOWER, J.C. (1971): "A general coefficient of similarity and some of its properties". *Biometric* Vol. 23 p. 623-637.
- HAYTER, R. (1975): "Farmers" crop decisions and the frost hazard in east-central Alberta". *Tijdschrift voor Econ. en Soc. Geografie* Vol. 66 p. 93-102.
- HENDERSON, H.V. y VELLEMAN, P.F. (1981): "Building multiple regression models interactively" *Biometrics* Vol. 37 p. 391-411.
- HEPPLE, L.W. (1981): "Spatial and temporal analysis: time series analysis" p. 92-96 en N. WRIGLEY y R.J. BENNETT *Quantitative geography*. London, Routledge Kegan and Paul.
- JOHNSTON, R.J. (1978): *Multivariate statistical analysis in Geography*. London, Longman.
- KMENTA, J. (1971) *Elements of econometrics*. New York, MacMillan.
- LABOVITZ, S. (1970): "The assignment of numbers to rank order categories". *American*

*Sociological Review* Vol. 35 p. 515–524.

LAZARFIELD, P.F. y HENRY, N.W. (1968): *Latent structure analysis*. New York, Houghton Mifflin.

LEIK, R.K. y GOVE, W.R. (1971): "Integrated approach to measuring association" p. 271–301 en H.L. COSTNER *Sociological methodology 1971*: San Francisco, Jossey Bass.

MATHER, P.M. (1976): *Computational methods of multivariate analysis in physical geography*. London, Wiley.

MATHER, P.M. (1981): "Factor analysis" p. 144–150 en N. WRIGLEY y R.J. BENNETT *Quantitative Geography*. London, Routledge and Kegan Paul.

MUELLER, J.H. y SCHUESSLER, K.F. (1961): *Statistical reasoning in sociology*. Boston, Houghton–Mifflin.

OPENSHAW, S. y TAYLOR, P.J. (1981): "The modifiable areal unit problem" p. 60–69 en N. WRIGLEY y R.J. BENNETT *Quantitative Geography*. London, Routledge Kegan and Paul.

POOLE, M.A. y O' FARRELL, P.N. (1971): "The assumptions of the linear regression model". *Transactions of the Institute of British Geographers* num. 52 p. 145–158.

PRINGLE, D.G. (1980): "Causal modelling: the Simon–Blalock approach" (CATMOG 27). Norwich, Geo Abstracts.

PHILIPS, L.D. (1973): *Bayesian statistics for social scientists*. London, Nelson.

RIPLEY, B.D. (1981): *Spatial statistics*. New York, Wiley.

SAVAGE, L.J. (1972): *The foundations of statistics*. New York, Dover.

TUKEY, J.W. (1977): *Exploratory data analysis*. Reading, Mass., Addison–Wesley.

UNWIN, D.J. (1978): "Quantitative and theoretical geography in the United Kingdom". *Area* Vol. 10 num. 5 p. 337–344.

WILSON, A.G. (1974): *Urban and regional models in geography and planning*. London, Wiley.

WILSON, A.G. (1971): "Catastrophe theory and bifurcation" p. 192–201 en N. WRIGLEY y R. J. BENNETT *Quantitative geography*. London, Routledge and Kegan Paul.

WRIGLEY, N. (1976): "Introduction to the use of logit models in geography" (CATMOG 10). Norwich, Geo Abstracts.

WRIGLEY, N. (1979): "Developments in the statistical analysis of categorical data". *Progress in Human Geography* num. 3 p. 315–355.

WRIGLEY, N. y BENNETT, R.J. (1981): *Quantitative Geography*. London, Routledge and Kegan Paul.

