

ANALISIS ESTADISTICO EXPLORATORIO Y CONFIRMATORIO EN GEOGRAFIA

Joaquín BOSQUE SENDRA
Departamento de Geografía
Universidad de Alcalá de Henares
Calle Colegios, 2
E-28801 Alcalá de Henares-Madrid

1) INTRODUCCION

Hace pocos años se han cumplido los veinte de la publicación del conocido libro Models in Geography (Chorley y Haggett, 1967), lo que ha provocado reuniones científicas y nuevas publicaciones (Macmillan, 1989), que han realizado un amplio balance sobre el período transcurrido y los cambios sufridos por la "Geografía cuantitativa" en este lapso de tiempo. En esta línea se sitúa el texto que sigue, donde por un lado se plantean algunos de los desarrollos, que nos han parecido mas importantes, en la metodología cuantitativa de la Geografía anglosajona y europea, y por otro se discuten nuevos métodos de análisis estadístico en relación a lo ocurrido en España en este mismo período.

2) NOVEDADES METODOLOGICAS RECIENTES EN LA GEOGRAFIA CUANTITATIVA

Una lectura rápida de algunos de estos trabajos que hacen balance de lo ocurrido en la Geografía cuantitativa anglosajona proporciona dos resultados bastante claros. En los últimos veinte años las dos novedades mas significativas son, por un lado el desarrollo de los denominados Sistemas de Información

Geográfica (SIG, o GIS en las siglas inglesas) y el planteamiento de un nuevo enfoque del propio método científico derivado del desarrollo de los modelos matemáticos dinámicos.

El primer tema es tratado en otras secciones de este volumen y por ello no va a recibir por mas tiempo nuestra atención, únicamente indicar su gran importancia práctica y el amplísimo volumen de publicaciones recientes que se le están dedicando en los últimos tiempos, numerosos textos, manuales y tratados (Bracken y Webster, 1990; Burrough, 1987), nuevas revistas: International Journal of GIS, de la editorial Taylor and Francis, y varios congresos científicos anuales, el mas reciente e importante para Europa la I Conferencia europea sobre Sistemas de Información Geográfica, celebrada en Amsterdam en abril de 1990.

El segundo tema planteado: los modelos matemáticos dinámicos y un nuevo enfoque del método científico, nos parece que es de un interés excepcional, por lo que supone de novedad no solo técnica sino mas aun metodológica e, incluso, con ciertas implicaciones conceptuales de importancia por ejemplo al eliminar o disminuir la tradicional distinción entre una Geografía nomotética y otra de carácter "idiográfico". Por ello lo vamos analizar con cierto cuidado.

3) FASES DEL MÉTODO CIENTIFICO CLASICO

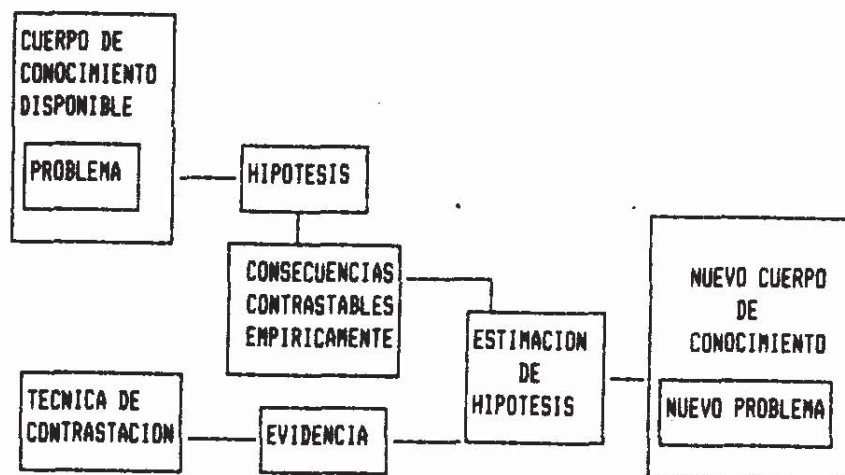
Existe una amplia discusión sobre las características del método científico y sobre todo de su aplicación a la Geografía y a otras ciencias sociales. Los trabajos de Bunge (1980, 1981 y 1985) contienen un buen resumen de las características del método científico. La figura 1 recoge un esquema (tomado de Bunge, 1981) de cual podría ser el funcionamiento

general del método científico en Geografía.

En lo esencial este esquema se puede describir en los siguientes términos:

1ª Fase de la elaboración de teorías: A partir del cuerpo de conocimientos existente en un momento dado, se plantean problemas sin resolver, a los que se enfrentan los científicos. Por lo tanto la Ciencia se puede entender como una empresa de resolución de problemas (Laudan, 1986). Frente a estos problemas el científico debe elaborar **hipótesis y teorías** que den cuenta del problema en los términos de lo ya conocido; las teorías, por su parte tienen que cumplir unos requisitos mínimos para ser válidas y adecuadas: "relacionar muchos hechos separados (...), sugerir nuevas relaciones (...), permitir hacer predicciones comprobables por la experiencia (...), ser la más simple posible en el sentido de requerir menos hipótesis o supuestos básicos (...), las hipótesis deben ser plausibles, aunque no estén sujetas inmediatamente a ensayo, y la teoría en conjunto no debe estar en conflicto con las ideas en boga (...), ser lo suficientemente flexible para desarrollarse y sufrir las modificaciones precisas" (Holton, 1976, p. 46-50).

Figura 1



Por otra parte, las teorías científicas se deben expresar en forma clara, precisa y comparable con la información observada (Principio de la intersubjetividad del conocimiento científico).

2º Fase de la obtención de datos y del contraste de la teoría: Al mismo tiempo, los mismos conocimientos existentes proporcionan de modo directo (las técnicas analíticas ya existentes) o indirecto (se pueden crear tales técnicas a partir de lo ya conocido), una serie de procedimientos de obtención y análisis de los datos reales. Todo ello facilita un método para contrastar las teorías con la realidad. En esta fase (comprobación de hipótesis) se comparan las consecuencias precisas de nuestras teorías e hipótesis con los datos reales obtenidos por los procedimientos de observación, experimentación y medida, y de acuerdo con ello se aceptará o rechazará la hipótesis formulada (aunque en otras ocasiones lo que se debe revisar son los propios datos reales, o, mejor dicho, los procedimientos que hemos utilizado para su obtención).

De esta manera habremos aumentado nuestro caudal de conocimientos, añadiendo una nueva teoría a las aceptadas en ese momento o eliminando por incorrecta unas técnicas y unas teorías previamente existentes. Normalmente, cualquier avance en el conocimiento de un fenómeno incrementa los problemas sin resolver; precisamente una buena prueba de su interés científico es que una cuestión abra nuevos problemas antes no conocidos. De este modo el ciclo sigue indefinidamente.

4) EL NUEVO ENFOQUE DE LA BIFURCACION

El procedimiento de investigación descrito ha sido creado por los científicos naturales, físicos y

astrónomos sobre todo, y ha sido aplicado y explica muy bien dos tipos de fenómenos (Allen y otros, 1984a y b): los sistemas físicos muy **simples**, donde funcionan leyes de conservación rígidas y, por lo tanto, es fácil la "predicción" de situaciones futuras, y, por otro lado, los sistemas físicos que estén próximos al **equilibrio** termodinámico, es decir de máximo desorden. En este caso la evolución temporal sólo conduce a la eliminación de estructuras complejas, a una situación de mayor equilibrio.

Ninguna de estas dos características está de acuerdo con lo que es la evolución de los fenómenos sociales ya que éstos se caracterizan mejor por su carácter complejo, evolutivo y jerárquico. Lo mismo ocurre con los fenómenos biológicos, por lo que en ocasiones se ha defendido igualmente la inutilidad del empleo del método científico clásico para el estudio de sistemas biológicos.

Los fenómenos sociales son complejos por poseer **siempre** una estructura interna (Aracil, 1986), a diferencia de multitud de hechos físicos de los que se puede asumir no poseen estructura interna discernible y solo se trata de conocer su dinámica e interacción con otros hechos igualmente simples. Los elementos internos componentes de cualquier hecho social están conectados mediante una estructura de relaciones causa-efecto, relaciones causales que pueden formar bucles de realimentación de diversos tipos.

En segundo lugar, los fenómenos sociales son evolutivos por que: "a base de cosas y sistemas de cosas previos, y como resultado de su actividad, se forman continuamente nuevos sistemas de cosas, que poseen propiedades emergentes, es decir que no son reducibles a las propiedades de sus componentes" (Quintanilla, 1981, pp. 23).

Finalmente los fenómenos sociales son jerárquicos ya que existen en ellos: "distintas capas o niveles de organización -los niveles resolutivos- cada uno de los cuales mantiene relaciones con sus colaterales -superior e inferior- pero de modo que a esta relación no se le puede considerar de dependencia unidireccional" (Aracil, 1986, p. 212).

Todas estas características no se dan, o al menos no con tanta intensidad, en los fenómenos físicos usualmente estudiados por las ciencias "duras".

El método científico clásico (que también ha sido llamado enfoque "newtoniano" o mecanicista de la ciencia) ha sido empleado, no obstante, con bastante éxito en el estudio de los hechos sociales (que eran considerados, por lo tanto, como simples y en equilibrio) y ha permitido llegar a formular teorías y modelos de gran utilidad, aunque no es el momento de enumerar la gran cantidad de teorías, hipótesis, conceptos y técnicas desarrolladas desde este punto de vista, ni tampoco las numerosas aplicaciones prácticas deducidas de estos trabajos, únicamente queremos mencionar algún libro donde se encuentran resultados de interés: por ejemplo, Amedeo y Golledge (1975) o Hagget y otros (1977). No obstante, siempre ha podido criticarse la limitada precisión de estas teorías y modelos y su reducida capacidad predictiva, generalmente mas bien meras extrapolaciones estadísticas con grandes probabilidades de resultar equivocadas. Esta ineficiencia relativa del método científico en su aplicación a los hechos sociales esta producida por la inadecuada concepción de la realidad social como similar a la realidad que estudia la Física, y no por las propias características epistemológicas del método científico: formulación de teorías y contraste con la realidad, que son válidas igualmente para el

estudio de los hechos sociales.

En los últimos años está surgiendo una alternativa a esta aplicación demasiado literal del método científico clásico a los hechos sociales y humanos. Procede de diversas fuentes, pero casi todas son desarrollos teóricos y metodológicos de las ciencias experimentales. Por un lado, de la recientemente creada (años 40-50) Teoría General de sistemas (Bertalanffy, 1976; Aracil, 1986) y, por otra parte, de la Física (Haken, 1986), en concreto de la Termodinámica de los procesos irreversibles, y en especial de los trabajos del premio Nobel de Química Ilya Prigogine (Prigogine, 1983; García Velarde, 1982). Prigogine analizaba sistemas físicos inicialmente homogéneos, "abiertos" (es decir que realizan intercambios de materia y energía con el medio ambiente) y situados lejos del equilibrio termodinámico, en ellos se producen espontáneamente fluctuaciones aleatorias en su composición que pueden conducir a que el sistema evolucione hacia una nueva posición de equilibrio, caracterizada esta vez por la falta de homogeneidad en su composición (Aracil, 1986, p. 173). A este proceso de generación de un orden nuevo en el sistema (de base espacial, pues las distintas zonas del sistema tienen ahora diferente composición) Prigogine lo ha denominado **orden mediante fluctuaciones**. Una característica importante que deben presentar los procesos que den origen a este tipo de evolución es la no linealidad y la existencia de bucles de realimentación en las relaciones entre sus elementos componentes. Ambas circunstancias son, precisamente, definitorias de los hechos sociales, por lo que es necesario pensar en la existencia de este tipo de situaciones en las Ciencias sociales.

Esta evolución es controlada por dos tipos de mecanismos, uno la fluctuación aleatoria inicial, el otro las leyes deterministas que guían el proceso, tal tipo de

mecanismos conducen a la necesidad de introducir la Historia de un fenómeno para llegar a entender su situación actual (a diferencia de los procesos de la Física clásica donde el tiempo era reversible), ya que en su evolución han surgido puntos de inestabilidad y de bifurcación en su trayectoria evolutiva, puntos ligados a fluctuaciones aleatorias y por lo tanto impredecibles. Esto impide una extrapolación determinista de su estado actual. Al mismo tiempo suponen admitir como acertadas las críticas derivadas de un planteamiento "idiográfico" historicista, que subrayaban la importancia central de la Historia previa para entender la situación actual de un fenómeno social.

Las formas de organización que son resultado de estos procesos se denominan "estructuras disipativas" (por ser estructuras que disipan, gastan energía para su mantenimiento) (García Velarde, 1982; Gea Banaclocha y García Velarde, 1981), que solo pueden existir en organizaciones que mantienen intercambios de materia y energía con el medio ambiente a través de sus fronteras, lo que les posibilita la ya citada realización de procesos de autoorganización espontáneos. Aunque existe una importante discusión sobre la migración sin problemas de estos planteamientos, con origen en la Física y la Termodinámica, a las Ciencias sociales, la mayoría de los autores coinciden en la utilidad práctica de ello y en la existencia de mecanismos sociales semejantes a los que señalan en la Termodinámica (Mosekilde, Aracil y Allen, 1988, p. 20).

Es en este tipo de organizaciones, las **estructuras disipativas**, donde se producen las fluctuaciones locales que pueden ser amplificadas por la estructura global, lo que desemboca en trayectorias evolutivas muy diferentes para objetos inicialmente semejantes. En este enfoque la historia previa del comportamiento del sistema es esencial para explicar su situación presente

(Pumain, 1989, p. 157). Esto constituye una limitación estructural a la capacidad predictiva del método científico, que ahora sólo podrá delimitar el conjunto de posibilidades futuras que el sistema puede adoptar y el sentido general de su evolución posterior: ascendente, en disminución, cíclico, etc. Es decir solo existe una capacidad de predicción cualitativa y no cuantitativa (Aracil, 1986), sin poder especificar en cual de todas las posibles situaciones estará exactamente en el futuro, ya que ello depende de pequeñas fluctuaciones locales impredecibles. En gran medida estas complicaciones vienen dadas, como ya hemos indicado, por la existencia de relaciones no lineales y de bucles de realimentación entre los distintos componentes de los sistemas estudiados, los cuales forman parte esencial de una técnica de simulación, La Dinámica de Sistemas (Bosque Sendra, 1988) que resulta una buena herramienta para tratar estas cuestiones.

Todas estas ideas y los procedimientos analíticos desarrollados en torno a ellos: tales como a) la teoría matemática de Catástrofes: Thom, (1987); Rodríguez Illera (1982); Woodcock y Davis (1986), Saunders (1983) y Wilson (1981) b) los modelos dinámicos de los fenómenos: Clarke y Wilson, 1983; Pumain y otros (1984); Lombardo y Rabino (1984); Culling (1987) y c) la geometría fractal de la naturaleza: (Mandelbrot, 1987) y de los hechos geográficos (Longley y Batty, 1986), se pueden aplicar a otro tipo de sistemas complejos, como los de carácter biológico o social. De este modo, la simplificación anteriormente mencionada: considerar como sistemas simples y en equilibrio a los sistemas sociales y geográficos desaparece.

Tenemos por lo tanto un desarrollo del método científico, el denominado **enfoque de la bifurcación o del orden a partir del caos o de las fluctuaciones**, que puede ser de enorme utilidad en el estudio científico

de los hechos sociales y por lo tanto en el desarrollo de una Geografía nomotética y científica renovada, que al mismo tiempo integre algunos de los elementos conceptuales y metodológicos tradicionalmente planteados por la escuela "idiográfica" o historicista de la Geografía.

Las implicaciones inmediatas de este profundo cambio metodológico son las modificaciones que es necesario realizar en los procedimientos utilizados en las dos fases que hemos diferenciado en el método científico: a) en el momento de la elaboración de teorías y modelos, resulta imprescindible el empleo de la **simulación por ordenador**, aquí resulta de gran interés la ya mencionada Dinámica de sistemas (Martínez Vicente, 1989; Chamussy, 1989) y b) en la fase de contraste de teorías, se tienen que producir cambios en la manera de medir los datos observados y en los propios métodos de análisis de datos, sobre esta cuestión veremos algunos planteamientos al tratar de lo ocurrido en la Geografía cuantitativa española en los últimos años,

Desde el punto de vista de las aplicaciones prácticas de estos planteamientos ha sido, hasta el momento, en el estudio de los temas urbanos donde mas ampliamente se han utilizado estos nuevos métodos. Y se han desarrollado con ello modelos originales y resultados de gran interés. Varias líneas diferentes, aunque convergentes, se pueden señalar. En primer lugar, el modelo matemático de la estructura urbana desarrollado por Forrester (1969), uno de los mas clásicos sobre el tema; Alfeld y Graham (1976) llevaron a cabo un análisis divulgativo de gran interés didáctico y recientemente Aracil y sus colaboradores están desarrollando el denominado "Análisis estructural" o "Análisis cualitativo" de su organización intentando determinar la existencia de inestabilidades y "atractores"

que pueden llegar a producir cambios cualitativos (Mosekilde, Aracil, Allen, 1988; Aracil y Toro, 1984; Toro y Aracil, 1988; Aracil, 1983; Aracil, Ponce y Freire, 1985), con ello añaden un nuevo elemento técnico de gran importancia a la metodología de la Dinámica de sistemas creada por Forrester. En segundo lugar resultan de gran importancia los modelos urbanos desarrollados por P. Allen, inspirándose directamente en las ideas de Prigogine, uno para los sistemas de ciudades, basado en los conceptos de la teoría del lugar central, planteada desde un punto de vista dinámico (Allen y otros, 1984a; Allen y Sanglier, 1979), el otro para la organización intraurbana (Allen y otros, 1984a y b), este último ha sido desarrollado y probado en ciudades francesas (Pumain, Saint-Julien, Sanders, 1984; Saint-Julien, 1989) e italianas (Lombardo y Rabino, 1984; Bertuglia y otros, 1985). Un tercer lugar lo ocupa la línea de trabajo desarrollada por Allan Wilson de la Universidad inglesa de Leeds que ha creado su propio planteamiento para los modelos dinámicos urbanos, con un extenso grupo de aplicaciones, un balance de ellas se puede encontrar en Clarke y Wilson (1983). Por último se deben mencionar los trabajos de Dendrinós que aplican las ideas del modelo Volterra-Lotka sobre la competencia interespecífica a la explicación del crecimiento relativo de un área metropolitana, con unos curiosos resultados empíricos relativos a la ciudad de Madrid (Dendrinós, 1984; Dendrinós y Mullaluy, 1985).

5) EL ANALISIS ESTADISTICO Y LA GEOGRAFIA CUANTITATIVA ESPAÑOLA EN LOS ULTIMOS AÑOS

En la Geografía cuantitativa de nuestro país, por el contrario, las cosas parecen ir por otros derroteros. Así un estudio de las Actas de los tres coloquios de Geografía de este carácter ya celebrados (Oviedo, 1983;

Madrid, 1986 y Cáceres, 1988) nos muestra que todavía se mantiene un predominio casi absoluto de las aplicaciones estadísticas y de análisis de datos, en especial entre las comunicaciones a ellos presentadas. Sin embargo, la formulación de modelos matemáticos tiene un papel mucho más secundario o es casi inexistente. Por ello, parece oportuno tratar en este texto algunas nuevas posibilidades y técnicas de análisis de datos, que nos parecen están mejor adaptadas a las necesidades peculiares de los geógrafos y de los datos espaciales, y por otra parte pueden ser una buena respuesta a los cambios introducidos en este tema por el nuevo enfoque del método científico que hemos tratado en el apartado previo. Nos referimos al denominado "análisis exploratorio de datos" (Hartwing y Dearing, 1979) y a los procedimientos de análisis confirmatorio basados en el remuestreo de la información observada, como el denominado "bootstrap" (Diaconis y Efron, 1983).

La Estadística que podemos llamar clásica, dividida en dos ramas, la descriptiva y la inferencial, se planteaba como objetivo fundamental "confrontar hipótesis con una imagen dada de la realidad" (Grupo CHADULE, 1980, p. 14). Para ello se realizan dos fases de trabajo, en primer lugar se describen y sintetizan los datos observados de manera que se les prepara para poder contrastar sus características generales (media, desviación típica, etc) con las que se derivan de la hipótesis a contrastar. Esto se lleva a cabo en la segunda fase la inferencial, usando los diversos tipos de test o pruebas estadísticas disponibles.

Este planteamiento tropieza con varias dificultades cuando se aplica a la Geografía. La primera y principal es que habitualmente el geógrafo no tiene formulada una hipótesis previa que se deba contrastar y que determine los datos a recoger y las pruebas o test

estadísticos a realizar. En segundo lugar, la Estadística clásica realiza dos suposiciones bastante fuertes sobre los datos a la hora de desarrollar sus métodos de análisis y de contraste de hipótesis. Por un lado se considera que los datos recogidos, y que se están describiendo, forman una muestra representativa de una población mucho mayor que no podemos observar en su totalidad. Por otra parte, se admite que, en general, la distribución de frecuencias de la variable estudiada sigue la ley normal o de Gauss en la población.

Ambas suposiciones son discutibles en Geografía. La distinción entre muestra y población no está claro que tenga sentido en muchas situaciones geográficas (por ejemplo, en los datos censales recogidas exhaustivamente) y, si lo tiene, es en un modo muy diferente al que considera la estadística clásica, a través del problema de la unidad espacial modificable (Bosque Sendra y otros, 1986), pues una concreta subdivisión del espacio se puede considerar una muestra de tamaño uno escogida de entre un conjunto infinito o muy grande de subdivisiones posibles del espacio. Pero aun admitiendo la validez de la distinción entre muestra y población lo que es todavía mas discutible es que las poblaciones geográficas se adaptan a la ley de Gauss en su distribución de frecuencias. Por todo ello las aplicaciones de la Estadística clásica a la Geografía están plagadas de problemas y de dificultades. Es necesario empezar a considerar alguna otra alternativa, y una buena ayuda en este sentido puede ser el Análisis Exploratorio y el Análisis Confirmatorio de Datos.

6) EL ANALISIS ESTADISTICO EXPLORATORIO (EDA) Y SU EMPLEO EN LA GEOGRAFIA

El EDA es un nuevo planteamiento para el

análisis de datos desarrollado, en especial, por el estadístico americano J.W. Tukey (1977). Un amplio conjunto de discípulos ha colaborado en su desarrollo: Hoaglin y Mosteller (1983); Hoaglin, Mosteller y Tukey (1985), Velleman y Hoaglin (1981). Mas recientemente se ha discutido sobre su interés en Geografía: Sibley (s.f. y 1984), Cox y Jones (1981), Anderson y Cox (1978); insistiendo en especial en su interés para la enseñanza de los métodos estadísticos a los estudiantes de Geografía: Cox y Anderson (1978), Bradshaw (1986); Bosque Sendra (1986). En España se ha empezado también a aplicar en otras ciencias sociales: Sierra Bravo (1985).

Las ideas básicas del EDA son, en resumen, dos. En primer lugar, el escepticismo ante las medidas estadísticas que intentan resumir una variable, pues aunque por un lado nos puedan indicar algo de esa variable, por otra parte, siempre nos ocultan algo de ella, resultando, en alguna medida, engañosas. Esto ocurre en especial cuando los datos estudiados no tienen una distribución de frecuencias gaussiana. El EDA no parte de esta suposición, existente en la Estadística clásica, admitir de entrada una distribución de frecuencias gaussiana y a partir de ello crear las medidas estadísticas, como es el caso de la media aritmética. En el EDA las medidas descriptivas de una variable tienen que servir sea cual sea su distribución de frecuencias.

La otra idea básica es la de apertura de ideas y actitudes, el análisis de datos no se plantea únicamente para contrastar una hipótesis previamente formulado. Su objetivo es hacer aparecer las "estructuras" que están embebidas en los datos observados, por ello los procedimientos deben ser capaces de eliminar el "ruido" que enmascaran estas estructuras subyacentes en la información.

Partiendo de estos principios fundamentales se han elaborado una amplia serie de métodos analíticos que complementan o sustituyen a los clásicos, capaces, por lo tanto, de realizar una descripción univariada, bivariada o multivariada de los datos, y de analizar distintas organizaciones de la información: listas de valores, series temporales o tablas de contingencia. A continuación vamos a comentar algunos de estos procedimientos insistiendo sobre todo en lo que los separa de los métodos mas tradicionales.

Desde un punto de vista técnico el EDA se caracteriza por su insistencia en el empleo de procedimientos analíticos y descriptivos de carácter gráfico o semigráfico (Cleveland, 1985; Chambers y otros, 1983). La razón es que tales métodos se consideran mas capaces de mostrar todas las particularidades de la variable estudiada, ocultando menos aspectos de ella y facilitando descubrir las "estructuras" existentes dentro de sus valores. En esto también coincide muy bien el EDA con las tradiciones geográficas de la "graficidad" y de su importancia como medio de comunicación y análisis de la información. En este sentido es muy útil comparar el típico histograma de frecuencias con el denominado "Gráfico tallo y hojas", la alternativa del EDA al estudio y representación de la distribución de frecuencias de una variable.

Figura 2

Histograma y Tallo y hojas: votaciones al PSOE en la ciudad de Granada (165 secciones)

Distribution & Histogram

variable: PSOE

Bin	Lower	Upper	Count	Prct	Total	Prct	Histogram
1	7.564297	12.01254	3	1.8	3	1.8	###
2	12.01254	16.46079	10	6.1	13	7.9	#####
3	16.46079	20.90904	18	10.9	31	18.8	#####
4	20.90904	25.35729	11	6.7	42	25.5	#####
5	25.35729	29.80554	10	6.1	52	31.5	#####
6	29.80554	34.25379	12	7.3	64	38.8	#####
7	34.25379	38.70203	11	6.7	75	45.5	#####
8	38.70203	43.15028	7	4.2	82	49.7	#####
9	43.15028	47.59853	10	6.1	92	55.8	#####
10	47.59853	52.04678	17	10.3	109	66.1	#####
11	52.04678	56.49503	18	10.9	127	77.0	#####
12	56.49503	60.94328	10	6.1	137	83.0	#####
13	60.94328	65.39153	14	8.5	151	91.5	#####
14	65.39153	69.83978	5	3.0	156	94.5	#####
15	69.83978	74.28803	7	4.2	163	98.8	#####
16	74.28803	78.73627	0	0.0	163	98.8	
17	78.73627	83.18453	2	1.2	165	100.0	##

(Stem & Leaf Display)

Depth	Stem	Leaves	Plot of PSOE
2	0	. #79	
9	1	#1223444	
28	.	#5556666677888899999	
42	2	#00011222233344	
52	.	#6666788999	
69	3	#11111233334444444	
76	.	#566777899	
(10)	4	#0111333444	
77	.	#55668888888999999	
61	5	#000112222333334444	
43	.	#55556678899	
32	6	#00001111223333444	
15	.	#557779	
9	7	#0001233	
2	.	#	
2	8	#23	

Unit = 1 Example: 1 #2 Represents 12

(Ambos gráficos han sido realizados con el programa de análisis estadístico NCSS)

Como se puede apreciar en la figura 2 la forma general de la distribución de frecuencias se observa igual en los dos gráficos, pero además el "gráfico tallo y hojas" muestra mas detalles dentro de cada tramo de la distribución, en realidad todos los valores numéricos de la variable están disponibles. Al mismo tiempo, nos informa de la llamada "profundidad" (depth), o distancia de cada tramo del gráfico al extremo mas cercano de la variable y, del tramo que contiene la mediana, en este caso el de 40-44%. En cualquier caso, es evidente en ambos gráficos la separación pronunciada de esta variable de una distribución de frecuencias gaussiana. Por otra parte el trazado de un gráfico "tallo y hojas" resulta tan sencillo de llevar a cabo, o incluso mas, que un histograma; se puede, por lo tanto, recomendar como el método de elección para una descripción manual y rápida de la distribución de frecuencias de una variable. Al mismo tiempo reorganiza los valores ordenandolos de menor a mayor.

Otra cuestión en la que insiste mucho el EDA es la denominada "resistencia" de los métodos de descripción, ante la existencia en la variable analizada de unos pocos casos con valores numéricos que se separan ampliamente del resto y que, por ello, pueden afectar excesivamente a los resultados de estos análisis, como es el caso del cálculo de la media aritmética y de la desviación típica. Utilizar la media aritmética para dar cuenta de la "tendencia central" o de la desviación típica para medir la variabilidad, puede producir en estos casos resultados muy equivocados y confusos. Por ello el EDA se inclina mas bien por el empleo de la mediana y de los cuartiles o percentiles, formando con ellos un cuadro, el denominado "resumen numérico" de una variable.

En el caso de la variable del ejemplo: votaciones al PSOE en la ciudad de Granada en 1987 (figura 3 y 4),

emplear la media para dar cuenta de su tendencia central nos puede confundir ampliamente, pues su distribución de frecuencias no es gaussiana. La mediana, por su parte, tiene un valor diferente y algo mayor y tampoco resulta completamente adecuada para indicarnos el valor más representativo de dicha variable. En realidad en estos casos no es posible usar una sola medida para obtener la tendencia central, es necesario considerar todos los valores y su distribución de frecuencias, tal y como lo muestra el gráfico tallo y hojas y el resumen numérico de la figura 3, en el cual podemos comprobar la posición altas y bajas de varios estadísticos (mediana y diversos percentiles), conocer sus diferencias y sus valores medios y de este modo tener una idea más completa y menos engañosa de su tendencia central y de su variabilidad.

**Figura 3. Resumen numérico de la variable:
votaciones al PSOE en la ciudad de Granada (165
secciones censales)**

n = 165						
Medida	Rango	Bajo		Alto	Media	Diferencia
-----	-----	-----		-----	-----	-----
Mediana	83		43.5		43.5	--
75% o 25%	41.5	24.6		55.06	39.83	30.46
87% o 12%	21	18		63	40.5	45
93% o 6%	10.5	15		67	41	52
96% o 3%	5.5	12.5		71.5	42	59
100% o 0%	1	7.5		83.1	45.3	75.6

Figura 4: Descripción numérica de una variable, votaciones al PSOE en la ciudad de Granada (165 secciones censales)

Variable: PSOE1

Mean - Average	41.51081	No. observations	165
Lower 95% c.i.limit	38.73148	No. missing values	0
Upper 95% c.i.limit	44.29013	Sum of frequencies	165
Adj sum of squares	53615.74	Sum of observations	6849.283
Standard deviation	18.08107	Std.error of mean	1.40761
Variance	326.9253	T-value for mean=0	29.49028
Coef. of variation	.4355751	T prob level	0.0000
Skewness	4.476526E-02	Kurtosis	-1.052518
Normality Test Value	0.928	Reject if >	1.033(10%) 1.052(5%)
100-ile (Maximum)	83.18453	90-ile	64.39394
75-ile	55.06446	10-ile	16.79065
50-ile (Median)	43.50649	Range	75.62022
25-ile	24.64539	75th-25th ile	30.41907
0-ile (Minimum)	7.564297		

(Realizada con el programa de análisis estadístico NCSS)

Figura 5: Gráfico en caja y gráfico lineal de la variable: votaciones al PSOE en la ciudad de Granada

```

-----
7.564297-----Line Plot / Box Plot-----83.1815
1 1 12 434325423421 5122144352221213 3232 6532525612114344231121 32 2 11
-----XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-----

```

El "resumen numérico" de una variable también se puede representar gráficamente, lo que facilita su comprensión y visualización rápida, para ello se obtiene el denominado "Gráfico en caja", ver figura 5. En ella se puede observar dos tipo de gráficos estadísticos diferentes, el llamado "gráfico lineal" que únicamente muestra la distribución de todos los valores dentro de una escala que oscila entre el valor mínimo y máximo de esa variable, indicando las variantes con mas frecuencia de casos y los huecos (intervalos sin ningún caso) existentes en la variable. En concreto aquí se observa como existe un caso muy alto que se separa ampliamente de los restantes dejando así un hueco en la distribución de frecuencias. Por su parte el "gráfico en caja" muestra la posición (en la misma escala entre el mínimo y el máximo) de varios importantes estadísticos: la mediana, los cuartiles del 75 y del 25% (que son precisamente las medianas de los dos partes en que divide a la variable su propia mediana) y la localización de los valores observados mas próximos a las denominadas "vallas" de la distribución de frecuencias. Las "vallas" son calculadas por el procedimiento de sumar al cuartil del 75% y restar al cuartil del 25%, respectivamente, la cantidad obtenida de multiplicar 1.5 por la diferencia entre los dos cuartiles citados. Los valores de la distribución que se encuentren por fuera de las citadas "vallas" son llamados "anómalos" (outliers), y son aquellos que están muy alejados del cuerpo principal de la variable. En este caso existen casos anómalos tanto por la parte inferior como la superior de la distribución.

Del mismo modo que se han desarrollado técnicas para la descripción exploratoria de una sola variable, también existen métodos para estudiar la relación entre dos variables de manera que se eviten los problemas derivados del uso de la regresión lineal clásica: falta de resistencia de la línea de ajuste y subordinación a la

distribución de Gauss.

Para ello se han planteado varios métodos de ajustar una recta al gráfico de dispersión de dos variables, destaca entre ellos la construcción de una recta de ajuste basada en las medianas y no en las medias de las variables, lo que aporta la resistencia buscada a los parámetros de la recta de ajuste. El procedimiento consiste esencialmente en lo siguiente:

1) Se realiza un gráfico de dispersión de las dos variables, cuyo ajuste deseamos conocer.

2) Se divide el intervalo de variación de la variable X en tres partes, de manera que en cada una de ellas se sitúen la tercera parte del número de las observaciones empleadas.

3) En cada una de las tres partes se obtienen los valores de las medianas de las dos variables, la X y la Y. De este modo se crea la tabla siguiente:

PORCION GRAFICO	MEDIANA X	MEDIANA Y
----- IZQUIERDA	----- MX(i)	----- MY(i)
CENTRO	MX(c)	MY(c)
DERECHA	MX(d)	MY(d)

4) La pendiente (m) de la recta de ajuste lineal entre la variable X y la Y se obtiene por la expresión:

$$m = \frac{MY(d) - MY(i)}{MX(d) - MY(i)}$$

5) La ordenada en el origen (b) de la recta se

calcula mediante la fórmula:

$$b = 0.33 ((MX(i) + MX(c) + MX(d)) + 0.33 (MY(i) + MY(c) + MY(d)))$$

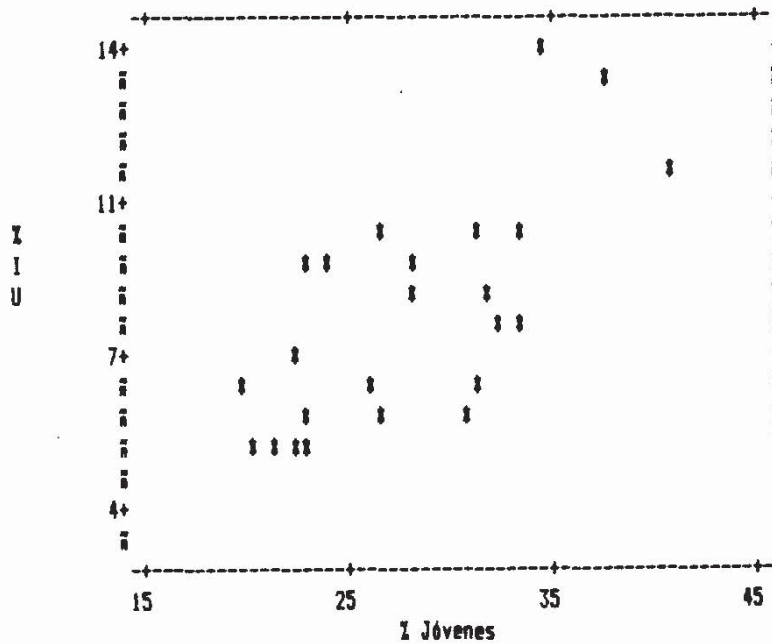
Esta recta de ajuste depende en muy poco grado de los valores extremos que puedan existir en el gráfico de dispersión, y sus parámetros son una función, sobre todo, de la parte central de los datos observados, por ello es mas representativa de la verdadera relación que existe entre las dos variables. Por el contrario, las rectas de regresión lineal están muy influidas por las posiciones y los valores numéricos de uno o dos casos que se separen fuertemente de los restantes. Por ello es recomendable obtener las dos rectas y cuando ambas coinciden es posible emplear la de regresión lineal sin ningún reparo sobre su verdadero significado. Este es el caso de lo que ocurre en la figura 6, que indica la relación entre las votaciones a la coalición electoral IU y el porcentaje de jóvenes en la población de cada uno de los veinticuatro barrios de la ciudad de Granada. Las ecuaciones de las dos rectas son las siguientes:

$$\text{Recta de ajuste resistente: } IU\% = -3.56 + 0.41 \cdot JOVENES\%$$

$$\text{Recta de regresión lineal: } IU\% = 0.07 + 0.29 \cdot JOVENES\%$$

Evidentemente los parámetros de ambas no son exactamente iguales, lo que ya muestra las diferencias en el procedimiento de ajuste utilizado en cada caso. No obstante, en este caso concreto las diferencias no son excesivas y, con las precauciones pertinentes, parece posible emplear la ecuación de regresión como una buena medida de la relación que existe entre las dos variables mencionadas.

Figura 6: Relación entre votos a IU y % de jóvenes



No son estas, evidentemente, las únicas posibilidades ofrecidas por el EDA para el análisis de los datos geográficos, a esto se puede añadir la importancia que alcanza en este enfoque la denominada "reexpresión" de los datos, de manera que se busque para cada variable la escala de medida (aritmética, logarítmica, etc) que sea mas adecuada para poner de manifiesto las estructuras subyacentes que sus valores pueden contener; se ha desarrollado para ello la denominada "familia de potencias" para llevar a cabo con facilidad la reexpresión de los datos mas conveniente a cada caso. Por ejemplo, el citado "resumen numérico" de una variable contiene información no solo sobre la tendencia central de la variable (a través del valor de la mediana) o de la variabilidad (las diferencias entre los cuartiles y los percentiles) además, también nos informa sobre la forma de la distribución de frecuencias de la variable

(observando la disposición de los valores de la columna de MEDIAS), y es, finalmente, una guía para usar la familia de potencias en búsqueda de la reexpresión mas adecuada para cada variable.

Por otra parte, con el EDA no solo se pueden analizar listas de valores como hasta este momento se ha tratado, igualmente existen métodos para describir una serie temporal, una tabla de contingencia o llevar a cabo análisis de la varianza de uno o mas factores

En resumen, el EDA constituye una magnífica herramienta para ser usada por los geógrafos en sus análisis de datos.

7) EL "BOOTSTRAP" Y OTROS MÉTODOS DE AUTOMUESTREO: POSIBILIDADES EN GEOGRAFIA

El Análisis exploratorio de datos nos permite explorar las características de los datos sin subordinar nuestros procedimientos a la suposición de que la variable tenga una distribución de frecuencias próxima a la de Gauss. Por su parte, el Análisis confirmatorio pretende realizar inferencias sobre la población sin tener que considerar la misma suposición de la distribución de Gauss. Los procedimientos inferenciales clásicos se basan precisamente en esta suposición, que no es admisible de modo general para los datos geográficos. La técnica que se emplea para poder realizar inferencias sin suponer la distribución de Gauss es el remuestreo de los datos observados (Hinkley, 1986).

El "bootstrap", y los otros métodos de automuestreo o remuestreo como el jackknife (Efron, 1982), se basa en la simulación por ordenador para el análisis de datos (Efron, 1982; Yang y Robinson, 1986, cap. 7). Con este método se puede estimar el sesgo, el

error típico y otras medidas del error de nuestras inferencias estadísticas sin necesidad de asumir ninguna hipótesis acerca de los datos. La idea básica es simular repeticiones de los datos y observar la variación estadística de las inferencias buscadas (por ejemplo, los coeficientes de regresión o de correlación de un análisis de regresión múltiple). Con las simulaciones de los datos se obtienen una amplia serie de muestras aleatorias, entre 100 y 1000 habitualmente, (remuestreadas de los datos originales) y a partir de ellas se establecen los valores de el error típico, el sesgo y otras medidas de la calidad de nuestras inferencias.

La justificación del método es el hecho bien establecido de que en gran número de situaciones el "bootstrap" es casi equivalente a la verdadera repetición de la observación de los datos con errores aleatorios reales, en cuyo caso, la varianza y el error típico calculados serían, respectivamente, la verdadera varianza y el verdadero error típico.

Este procedimiento general puede ser aplicado a cualquier tipo de análisis inferencial: obtención de la media y de la desviación típica de una población (Efron y Tibshirani, 1986); establecer la recta de regresión entre dos variables e incluso para realizar análisis multivariados: Análisis factorial por ejemplo (Diaconis y Efron, 1983). Evidentemente en cada caso será preciso aplicar un procedimiento ligeramente diferente aunque siempre basado en los mismos principios. En lo que sigue estudiamos como realizar un análisis de "bootstrap" para comprobar la validez y los errores de los coeficientes de una ecuación de regresión lineal múltiple (Folmer y Fischer, 1986).

El primer paso es estimar los errores aleatorios (e) de la ecuación de regresión múltiple para cada observación. Se estiman los parámetros de la ecuación

de regresión por el método de los mínimos cuadrados. A su vez, los errores aleatorios existentes en cada observación se estiman mediante los residuos, calculados como el valor de la variable dependiente (Y) observada menos el valor calculado (YCAL) a partir del ajuste de la ecuación, es decir:

$$e = Y - YCAL$$

La magnitud de la variación aleatoria, es decir la varianza de los errores aleatorios citados, debe ser constante. Esto se puede comprobar mediante una gráfica de los valores de las variables dependientes frente al valor estimado de Y (YCAL). Si los residuos son homogéneos pueden servir como una muestra para simular, a partir de ellos, los errores experimentales. La simulación de los verdaderos resultados (los que existen en la población y no en la muestra usada) se consigue utilizando el modelo de regresión como la verdadera relación en cada uno de los casos, a la cual se le suma el valor de uno de los residuos antes calculado muestreado aleatoriamente y con reemplazamiento.

De este modo se crea un conjunto de resultados simulados (designados mediante *) formados por:

- Y*, obtenida mediante la ecuación de regresión ajustada mas el residuo extraído **aleatoriamente**.
- Las distintas X o variables independientes utilizadas inicialmente para estimar la ecuación de regresión.

En cada conjunto de datos, por lo tanto, los valores de las Y de las mismas observaciones difieren en función de cual ha sido el residuo aleatorio que se le ha sumado. Ver en la figura 6 un ejemplo de la creación de uno de los cien conjuntos de datos simulados.

Figura 6: Generación de uno de los conjuntos de datos simulados

BARRIO	1	2	3	4	5	6	7	8
CERRILLO	68.76	57.27	7.99	10.50	62.96	5.00	6	7.15
MAZAGRANDE	82.47	74.45	15.75	6.81	82.54	-0.07	13	-0.93
ZAIDIN	50.47	63.01	10.35	12.70	53.97	-3.50	19	0.37
CERVANTES	45.34	50.51	0.31	15.17	37.67	7.67	17	7.14
SIERRA	51.40	60.67	9.54	13.04	51.11	0.37	2	3.90
RECOSIANO	9.41	56.59	9.95	21.23	9.04	0.37	2	5.81
SANNTON	21.92	55.09	10.05	20.20	23.19	-1.27	12	0.76
ANGUSTIAS	33.29	56.13	13.59	19.02	31.40	3.09	21	-0.42
REALEJO	31.40	64.26	15.12	17.20	39.90	-0.42	0	-1.59
RECONDIST	24.93	49.99	0.56	15.02	27.01	-2.00	14	-4.33
RECONDIST	30.16	40.25	9.23	13.94	31.43	-1.27	12	2.47
LATINO	17.64	50.33	10.93	17.09	20.52	-2.00	14	-1.20
BURBUSA	24.60	52.01	11.90	21.90	17.47	7.15	4	-0.07
SANPATIAG	32.00	50.52	10.00	19.33	36.33	-4.33	10	-2.00
CENTRO	36.40	55.70	15.30	17.00	32.50	3.90	5	-3.10
UNIVERSIDA	33.91	40.96	12.96	15.10	26.76	7.15	1	-4.16
CIUDA	56.21	54.05	11.65	11.49	49.06	7.15	1	7.67
SANFRANCIS	25.57	46.90	5.79	11.21	38.93	-13.36	24	-2.49
CALVOBOTEL	36.51	40.05	11.27	13.00	29.59	6.92	23	-3.51
ALBAICIN	50.90	60.51	14.97	13.02	54.07	-3.17	15	3.79
AREAL	44.05	50.53	15.76	11.74	43.60	0.37	3	3.09
ALHAMBAYAR	49.19	53.30	5.63	9.25	51.60	-2.49	10	-3.60
MONTIJO	59.01	60.63	5.45	11.20	51.06	7.15	1	6.93
CARTAJA	73.72	69.91	0.75	7.76	69.94	3.70	20	-13.36

1: Valor de la votación al PSOE "bootstrap", obtenida sumando a la columna 5 la columna 6. 2: Variable explicativa: MACGRAN. 3: Variable explicativa JUBILANZ. 4: Variable explicativa PATRONOZ. 5: Votación al PSOE calculada con la ecuación de regresión (PSOE = 1.186 MACGRAN - 0.072 JUBILANZ - 2.772 PATRONOZ + 13.703). 6: Residuo sumado a la columna 5. 7: Número aleatorio que indica que residuo de la columna 6 se suma a la columna 5. El conjunto de datos sometido de nuevo al cálculo de la ecuación de regresión esta formado por la columna 1 como variable dependiente y las columnas 2,3 y 4 como variables explicativas.

**Figura 7: Análisis de bootstrap
Relación entre votos al PSOE y estructura social**

VARIABLE	VALORES COEF. REG.	SESGO	ERROR TIPICO	COEF. REG. MENOS SESGO
MACGRAN	1.186	-1.162	2.331	2.347
JUBILANZ	-0.072	0.070	4.443	-0.142
PATRONOZ	-2.772	0.060	4.152	-2.833
ORDENADA EN EL ORIGEN	13.703	-2.367	150.854	16.070
CORRELACION MULTIPLE: R	0.919	-0.022	0.404	0.941

Cada conjunto de resultados simulados se somete al mismo análisis estadístico que la muestra original, es decir cálculo de la ecuación de regresión múltiple ajustada por el procedimiento de los mínimos cuadrados, y obtención de los coeficientes de regresión y de correlación. El proceso de simulación se lleva a cabo un cierto número de veces (entre 100 y 1000). De esta forma se simulan observaciones repetidas con los mismo casos originales, obteniendo una serie de valores "bootstrap" de los parámetros de la regresión y del coeficiente de correlación. A partir de ellos se puede estimar el error típico, el sesgo y otras medidas de la calidad de la estimación de Y. Asimismo se pueden hallar coeficientes de regresión corregidos del sesgo que introduce la selección de una muestra de un tamaño muy inferior al de la población. Por lo tanto, con esta técnica realizamos inferencias sobre la ecuación de regresión sin tener que suponer la existencia de una distribución de frecuencias gaussiana en la población.

Si tenemos, por ejemplo, N estimaciones simuladas de los coeficientes de regresión:

$$a^*_1, a^*_2, \dots, a^*_n$$

una sencilla aproximación al sesgo de esa estimación relativa al valor de a es:

$$\text{sesgo}(a^*_i) = N^{-1} \sum_{j=1}^N a^*_j - \hat{a}$$

Esta relación representa la estimación de "bootstrap" del sesgo de a relativo al verdadero valor de a (en la población). De modo análogo, una simple aproximación a la varianza de la estimación es:

$$\text{var}(a^*) = N^{-1} \sum (a^*_j - \text{media}(a^*))^2$$

cuya raíz cuadrada es el error típico "bootstrap" de a .

La distribución de frecuencias del error de la estimación, $a - \hat{a}$ se puede aproximar mediante un histograma de las N diferencias $a_i^* - \hat{a}$ (Efron y Tibshirani, 1986).

Con ello no solo tenemos una serie de valores de los coeficientes de regresión/correlación corregidos del sesgo sino también una medida del error típico de cada coeficiente de regresión.

Un aspecto importante del bootstrap es que, una vez definido el modelo de regresión, es puramente numérico. Lo único que se requiere es un programa de ordenador que permita realizar repetidamente análisis de regresión (capacidad de generar macros) y un mecanismo de muestreo aleatorio para seleccionar los residuos. Por ello este procedimiento de análisis se ha llamado "de uso intensivo del ordenador" ya que solo es posible realizarlo disponiendo de ordenadores que nos permitan repetir las estimaciones de un modo rápido y cómodo (Diaconis y Efron, 1983). En este caso concreto se ha elaborado un programa informático específico (BOOTREG.EXE) en lenguaje TURBO BASIC, diseñado para el análisis "bootstrap" de la regresión múltiple. Normalmente algunos otros programas estadísticos se pueden usar para preparar los datos y para el estudio posterior de las estimaciones "bootstrap".

Un parámetro sujeto a elección es el número (N) de observaciones simuladas. Puede ser en torno a 100 para el cálculo del sesgo y del error típico y en torno a 1000 para poder determinar los intervalos de confianza de una estimación.

Como muestra de sus posibilidades hemos

realizado un análisis de la relación existente entre la variable dependiente "Porcentaje de votos al PSOE en la ciudad de Granada sobre total de votantes, 1987: PSOE%", frente a un conjunto de variables explicativas, en concreto las tres siguientes:

- Porcentaje de población nacida en la ciudad sobre total de habitantes: NACGRA%.

- Porcentaje de población que solo ha realizado estudios primarios sobre total de habitantes: ANALFAB%.

- Porcentaje de patronos y autónomos sobre total de activos: PATRONO%.

Todas las variables están medidas en los veinticuatro barrios en que se divide la ciudad.

El objetivo es determinar la verdadera ecuación de relación lineal entre la variable dependiente y las variables explicativas, para ello calculamos la ecuación de relación estimada y la corregimos del sesgo existente entre la verdadera y la estimada a través del uso del bootstrap. Los resultados se muestran en la figura 6. Las distribuciones de frecuencia de algunos de los errores (la diferencia entre el valor "bootstrap" y el valor estimado) se muestran en las figuras 8 y 9. El signo de los sesgos mostrados en la figura 7 indica que variables tienden a ser subestimadas. En concreto la variable explicativa NACGRA% corregida del sesgo ve incrementada su influencia sobre la dependiente en una cifra importante. No obstante no se llega a modificar el sentido de la influencia de ninguna de las tres variables explicativas. Las distribuciones de frecuencias de los errores de la estimación muestran un alto grado de aplastamiento, el cual está relacionado al valor del sesgo y del error típico, por ello en la variable

JUBILAD% existen varios valores anómalos y muy separados del resto, lo que muestra que por azar la estimación puede ser bastante diferente a las usuales. El coeficiente de correlación corregido de sesgo aumenta en un casi dos o tres puntos lo cual es bastante significativo. En resumen, el "bootstrap" no solo nos modifica la ecuación de regresión además nos proporciona unas medidas de la calidad de cada estimación a través de los errores típicos, lo cual puede ser de utilidad para determinar cual sería el modelo de regresión mas adecuado, permitiendonos seleccionar las variables explicativas mejores.

8) CONCLUSIONES

Para terminar, y como conclusión general, insistir en la aparición, desde varios puntos de vista, de un conjunto de novedades no solo técnicas, como las aquí descritas: el Análisis exploratorio y confirmatorio de datos, también de orden mas profundo, conceptual y metodológico, como son el nuevo enfoque del método científico que hemos discutido en los primeros apartados de este texto. Nos parece que, aunque no parezca existir una clara relación entre los dos temas, ambos comparten dos cosas. Un planteamiento de continuidad de los planteamientos clásicos, continuidad en lo que tienen de perenne e importante, y una necesidad de renovación y, sobre todo, de crítica de algunos supuestos, a veces no reconocidos, que subyacen a los planteamientos clásicos.

Crítica que en general consiste en diferenciar mas claramente los hechos sociales de los fenómenos naturales en diversos aspectos. Así por ejemplo, en un caso (el del enfoque de la bifurcación) se ha referido a la inadecuada consideración de los hechos sociales y geográficos como iguales a los puramente físicos desde el punto de su estructura interna. En otro, a la existencia de la distribución de Gauss, muy evidente en muchas situaciones de la Naturaleza, pero no existente en muchos fenómenos sociales.

Por lo tanto, la Geografía como cualquier otra ciencia social puede, e incluso debe, importar ideas y métodos de las ciencias experimentales, pero también necesita discutir y modificar tales ideas y métodos en lo que sea necesario para acomodarlas a las necesidades específicas del análisis de los hechos sociales.

Figura 8: Distribución de frecuencia de los errores de los coeficientes de regresión bootstrap de la variable: JUBILAD%

(Stem & Leaf Display)

Depth	Stem	Leaves	Plot of ERRJUBIL
	LOW	-132,-125,-104	
4	-8	66	
6	-7	660	
8	-6	672	
10	-5	631	
11	-4	61	
13	-3	681	
22	-2	6965333211	
31	-1	6876432221	
41	-0	69876532220	
(11)	0	600112456779	
48	1	6111133578889	
36	2	6123459	
30	3	6112223479	
21	4	6125799	
15	5	613699	
10	6	612678	
5	7	612	
3	8	69	
2	9	67	
	HIGH	6138	

Figura 9: Distribución de frecuencia de los errores de los coeficientes de regresión bootstrap de la variable: PATRONO%

(Stem & Leaf Display)

Depth	Stem	Leaves	Plot of ERRPATRO
	LOW	6-142	
5	-6	67610	
8	-5	6332	
14	-4	6883110	
21	-3	68652220	
25	-2	65330	
37	-1	6988776554100	
45	-0	699876420	
(9)	0	6233466779	
46	1	60123499	
39	2	6000225557799	
27	3	6155778	
21	4	60577	
17	5	611225568	
9	6	63788	
5	7	64	
4	8	645	
2	9	63	
1	10	61	

Unit = .01 Example: 1 62 Represents 0.12 Missing values: 1

9) BIBLIOGRAFIA

ALFELD, L.E. y GRAHAM, A.K. (1976): Introduction to Urban Dynamics Cambridge, Mass., Wright-Allen Press

ALLEN, D.M. y SANGLIER, M. (1979): "A dynamic model of growth in a central place system" Geogr. Analysis, 11, n. 3, 256-272.

ALLEN, P. y otros (1984a): "L'autoorganisation des systemes urbaines" en Y. GUERMOND (editor), Analyse de Systeme en Géographie Lyon, Presses Universitaires de Lyon, 201-254.

ALLEN, P.M. y otros (1984b): "Towards a new synthesis in the modelling of evolving complex systems" Envir. and Plan. B, 11, 65-84.

AMEDEO, D. y GOLLEDGE, R. (1975): An Introduction to scientific reasoning in Geography Nueva York, J. Wiley

ANDERSON, E.W. y COX, N.J. (1978): "A comparison of different instruments for measuring soil creep" Catena, 5, pp. 81-93

ARACIL, J. (1983): Introducción a la dinámica de sistemas Madrid, Alianza Editorial, 303 pp.

ARACIL, J. (1984): "Qualitative Analysis and Bifurcation in System Dynamics Models" IEEE Trans. on Syst. Man and Cybern., vol. SMC-14, n. 4, 688-696.

ARACIL, J. (1986): Máquinas, sistemas y modelos. Un ensayo de sistémica. Madrid, Tecnos, 282 pp..

ARACIL, J., PONCE, E., FREIRE, E. (1985): "Order through Fluctuations, and Systems Dynamics Models" Envir. and Plan. B, 12, 103-112.

ARACIL, J. Y TORO, M. (1984): "A case study of qualitative change in system dynamics" Int. J. Systems Sci., vol. 15, n. 6., 575-599.

BERTALANFFY, L. vON (1976): Teoría general de sistemas Mexico (edición original de 1968), F.C.E.,

BERTUGLIA y otros (1985): "Residencial location and urban dynamics processes: theoretical aspects and modelling" Budapest, 25º European R.S.A. Congress.

BOSQUE SENDRA, J. (1986): "El análisis estadístico exploratorio y la enseñanza de de las técnicas cuantitativas en Geografía" en Métodos cuantitativos en Geografía: enseñanza, investigación y planeamiento Madrid, A.G.E.

BOSQUE SENDRA, J. (1988): "Simulación por ordenador de procesos sociales" en Aplicaciones de la informática en Geografía y en ciencias sociales Madrid, Síntesis, pp. 237-270.

BOSQUE SENDRA, J. y otros (1986): "Algunos problemas metodológicos de las técnicas cuantitativas en Geografía humana" en La Geografía cuantitativa. Concepto y métodos. Oviedo, Pub. Universidad de Oviedo, 55-074.

BRACKEN, I. y WEBSTER, C. (1990): Information technology in Geography and planning. Including Principles of GIS Londres, Routledge.

BRADSHAW, R. (1986): "El futuro de la

Geografía cuantitativa" en La Geografía teórica y cuantitativa: concepto y métodos. Oviedo, Publicaciones de la Universidad de Oviedo.

BUNGE, M. (1980): Epistemología. Barcelona, Ariel, 275 pp.

BUNGE, M. (1981): La investigación científica. Barcelona. Ariel.

BUNGE, M. (1985): Racionalidad y realismo. Madrid, Alianza editorial,

BURROUGH, P. A. (1987) Principles of Geographical Information Systems for Land Resources Assessment. Oxford, Clarendon Press.

CLARKE, M. y WILSON, A.G. (1983) "The Dynamics of urban spatial structures: progress and problems" Journal of Regional Science, 23, 1, pp. 1-18

CLEVELAND, W.S. (1985): The elements of graphing data. Monterey, Cal., Wadsworth

COX, N. y ANDERSON, E. (1978): "Teaching geographical data analysis: problems and possible solutions" Journal of Geography in Higher Education, 2(2), pp. 29-37

COX, N.J. y JONES, K. (1981): "Exploratory data analysis" en N. WRIGLEY y R.J. BENNETT (editores), Quantitative Geography Londres. Routledge and Kegan Paul.

CULLING, W.E.H. (1987): "Equifinality: modern approaches to dynamical systems and their potential for geographical thought" Londres, Trans. Inst. British Geogra., N.S., 12, 57-72

CHAMBERS, J. M. y otros (1983): Graphical methods for data analysis Belmont, Calif., Wadsworth

CHAMUSSY, H. (1989): "La modelisation des systèmes dynamiques en Géographie. Problèmes théoriques et méthodologiques" Norba, VIII-IX, pp. 87-108.

CHORLEY, R.J. y HAGGETT, P. (Editores) (1967): Models in Geography. Londres, Methuen.

DENDRINOS, D.S. (1984): "Madrid's aggregate growth pattern: a note on the evidence regarding the urban Volterra-lotka model" Sistemi Urbani, 2, 237-246.

DENDRINOS, D.S. y MULLALUY, H. (1985): Urban Evolution. Studies in the Mathematical Ecology of Cities Londres, Oxford Univ. Press, .

DIACONIS, P. y EFRON, B. (1983): "Métodos estadísticos intensivos por ordenador" Investigación y ciencia, 82, pp. 70-83

EFRON, B. (1982): The Jackknife, the Bootstrap and other Resampling Plans Philadelphia, Society for Industrial and Applied Mathematics.

EFRON, B. y TIBSHIRANI, R. (1986): "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy" Statistical Science, vol. 1, 1, pp. 54-77.

FOLMER, H. y FISCHER, M.M. (1986): "Bootstrapping in spatial analysis" Besancon, Cahiers de Géographie de Besancon, 281-293

FORRESTER, J.W. (1969): Urban Dynamics Cambridge, Mass., MIT Press,

GARCIA VELARDE, M. (1982): "Estructuración y cooperatividad a partir del desorden" en Actas del I Congreso de Teoría y Metodología de las ciencias Oviedo, Pentalfa, 87-94

GEA BANACLOCHE, J. y G. VELARDE, M. (1981): "Estructuras disipativas: Potenciales y catástrofes (2)" Oviedo, El Basilisco, n. 12, 14-18

GUERMOND, Y. (editor) (1984): Analyse de Systeme en Géographie Lyon, Press Univ. de Lyon, .

HAGGETT, P. y otros (1977): Locational Models Londres, Arnold,

HAKEN, H. (1986): Formulas del éxito en la Naturaleza Barcelona, Salvat,

HARTWING, F. y DEARING, B.E. (1979): Exploratory data analysis Londres. Sage University Paper.

HINKLEY, D.V. (1986): "A Review of Resampling Methods for Confidence Limits" Technical Report nº 35, Center for Statistical Sciences, The University of Texas, Austin.

HOAGLIN, D.C. y MOSTELLER, F. (1983): Understanding robust and exploratory data analysis Nueva York. J. Wiley.

HOAGLIN, D.C, MOSTELLER, F. y TUKEY, J.W. (1985): Exploring data, tables, trends, and shapes Nueva York, J. Wiley.

HOLTON, G. (1976): Introducción a los conceptos y teorías de las ciencias físicas Barcelona, Ed. Reverte,

LAUDAN, L. (1986): El progreso y sus problemas. Hacia una teoría del crecimiento científico Madrid, Encuentro Ediciones, 295 pp.

LOMBARDO, S.T. y RABINO, G.A. (1984): "Nonlinear dynamics models for spatial interaction. The results of some empirical experiments" Papers of Regional Science, vol. 55, 83-101

LONGLEY, P. y BATTY, M. (1986): "Measuring and simulating the structure and form of cartographic lines" Papers in Planning Research. Agosto,

MACMILLAN, B. (Editor) (1989): Remodelling Geography Londres, Basil Blackwell

MANDELBROT, B.B. (1984): "De los monstruos de Cantor y Peano a la geometría fractal de la naturaleza" en Pensar la matemática Barcelona, Tusquets, 111-138.

MARTINEZ VICENTE, S. (1989): "Dinámica de sistemas y planificación regional" Norba, VIII-IX, pp. 63-86

MOSEKILDE, E., ARACIL, J. y ALLEN, P. (1988): "Instabilities and chaos in nonlinear dynamic systems" System Dynamics Review, 4, 1-2, pp. 14-55

PRIGOGINE, I. (1983): ¿Tan sólo una ilusión? Una exploración del caos al orden Barcelona, Tusquets

PUMAIN, D. (Editores) (1989): "Spatial Dynamics and Urban Model" en J. Hauer, H. Timmermans y N. Wrigley: Urban Dynamics and Spatial Choice behaviour

Dordrecht, Kluwer, pp. 155-174

PUMAIN, D. y otros (1984): "Dynamics of spatial structure in french urban agglomerations" Papers of the Regional Science, vol. 55, 71-81.

PUMAIN, D., SAINT-JULIEN, Th. y SANDERS, L. (1984): "Vers une modelisation de la dynamique intra-urbaine" Paris, L'Espace Géographique, n. 2, 125-135.

PUMAIN, D., SAINT-JULIEN, Th. y VIGOUROUX, M. (1983): "Jouer de l'ordinateur sur un air urbain" Paris, Ann. de Geog., 511, 331-346.

PUMAIN, D., SAINT-JULIEN, Th y SANDERS, L. (1987): "Application of a Dynamic Urban model" Geographical Analysis, vol. 19, n. 2, 152-166.

QUINTANILLA, M.A. (1981): Fundamentos de Lógica y Teoría de la Ciencia Salamanca, Ed. Universidad de Salamanca,

RODRIGUEZ ILLERA, J.L. (1982): "Teoría de catastrofes y Ciencias sociales. Una entrevista a R. Thom" Oviedo, El Basilisco, n. 13, 70-73

SAINT-JULIEN, Th. (1989): "La modelización de la dinámica intraurbana aplicada al campo de algunas grandes ciudades francesas" Norba, VIII-IX, pp. 109-124

SAUNDERS, P.T. (1983): Una introducción a la teoría de catástrofes Madrid, Siglo XXI

SIBLEY, D. (S.F.): Spatial applications of exploratory data analysis. CATMOG, nº 49.

SIBLEY, D. (1984): "A robust analysis of a minority census: the distribution of Travelling people in England" Environment and Planning, A, 16, pp. 1279-1288

SIERRA BRAVO, R. (1985): "El análisis exploratorio y su aplicación a las tasas de natalidad, nupcialidad y mortalidad de las provincias españolas" Revista Internacional de Sociología, 43(4), pp. 635-665

THOM, R. (1987): Estabilidad estructural y morfogenesis Barcelona, GEDISA.

TORO, M. y ARACIL, J. (1988): "Qualitative analysis of system dynamics ecological models" System Dynamics Review, 4, 1-2, pp. 5-80

TUKEY, J.W. (1977): Exploratory Data Analysis Reading. Addison-Wesley.

VELLEMAN, P. y HOAGLIN, D. (1981): ABC of EDA Boston. Duxbury Press.

WILSON, A.G. (1981): Catastrophe Theory and Bifurcation: Applications to Urban and Regional Systems Londres, Croom Helm, 331

WOODCOCK, A. y DAVIS, M. (1986): Teoría de las catástrofes Madrid, Cátedra,

YANG, M.C.K. y ROBINSON, D.H. (1986): Understanding and learning Statistics by computer. World Scientific.