

# EL ANALISIS ESTADISTICO EXPLORATORIO Y LA ENSEÑANZA DE TECNICAS CUANTITATIVAS EN GEOGRAFIA

Joaquín Bosque Sendra  
(Universidad Complutense de Madrid)

## 1. INTRODUCCION

En la actualidad la enseñanza de Técnicas cuantitativas es un tema marginal dentro de la Geografía universitaria española (BOSQUE SENDRA y otros, 1983). Dentro de esta asignatura marginal parece un hecho general que los métodos estadísticos suponen una parte decisiva, y en muchos casos única, de su contenido. Por ello nos parece importante una discusión de la adecuación pedagógica de la enseñanza de estas técnicas estadísticas a los estudiantes de Geografía y el planteamiento de alternativas didácticas más adecuadas.

## 2. PROBLEMAS EN LA ENSEÑANZA DE LA ESTADISTICA EN GEOGRAFIA

Las dificultades de la enseñanza de los métodos estadísticos clásicos a los estudiantes de Geografía es una cuestión que ha sido debatida con amplitud en la Geografía anglosajona, muchos de los argumentos allí esgrimidos nos parecen de aplicación en nuestro país. (COX, N. y ANDERSON, E., 1978; COX, N., 1978; COX, N. y ANDERSON, E., 1980; EHRENBERG, A., 1979). Según estos autores los problemas más importantes serían esencialmente:

- a) En primer lugar, la falta de conocimientos matemáticos adecuados de los alumnos de Geografía, lo cual constituye una dificultad todavía más relevante en la universidad española. La gran mayoría de nuestros estudiantes de Geografía no han cursado ninguna asignatura de Matemáticas de nivel universitario, y por otro lado, sus estudios de Matemáticas en el bachillerato están totalmente olvidados (dada su lejanía temporal, normalmente has-

ta cuatro años) o, peor aún, supusieron un verdadero trauma por lo que se ha creado un recelo ante cualquier aspecto matemático que ahora dificulta mucho el aprendizaje de estas técnicas cuantitativas.

b) En segundo lugar, la falta de validez de la Estadística clásica (sobre todo la inferencial) en su aplicación a los datos geográficos. Este es un tema repetidamente discutido (ver por ejemplo un resumen en BOSQUE SENDRA, J. y otros, 1983 b) por lo que sólo recordamos los aspectos más importantes:

- La cuestión de la falta de definición precisa del individuo geográfico, que produce el denominado "problema de la unidad espacial modificable", fuente de indefiniciones gravísimas en los usualmente considerados muy precisos resultados de los métodos estadísticos.

- La autocorrelación espacial y la falta de "normalidad" gaussiana de los datos geográficos, que dificultan el uso de test estadísticos (HAGGETT, P. y otros, 1977).

c) En tercer lugar la poca relevancia geográfica de las aplicaciones más usuales de la Estadística. Esta es, en nuestra experiencia concreta como profesor de esta asignatura, la crítica más usual formulada por los alumnos. Los cuales piensan que gran parte de lo que aprenden en ella tiene poco que ver con los problemas más corrientes del resto de las asignaturas geográficas. Lo que en parte se origina por la enseñanza de los procedimientos estadísticos de comprobación de hipótesis y de los test de significación, que en una disciplina no experimental como la Geografía pierden gran parte de su interés, ya que la Geografía, como ciencia más centrada en la observación, necesita más bien métodos de descripción y análisis de datos.

### 3. UNA ALTERNATIVA AL EMPLEO DEL ANALISIS ESTADISTICO EN GEOGRAFIA

Todo lo anteriormente expuesto constituye un conjunto de razones que deberían llevarnos a un replanteamiento de la actual

insistencia exclusiva en los métodos estadísticos, sobre todo cuando ya existen alternativas muy interesantes, como el Análisis estadístico exploratorio de datos (EDA), que por un lado los puede sustituir con ciertas ventajas y por otro los puede complementar y enriquecer.

El EDA es un conjunto de procedimientos de análisis de datos desarrollados en los últimos veinte años por varios matemáticos y estadísticos, en especial J.W. Tukey de la Universidad de Princeton y de los Laboratorios Bell de los Estados Unidos, como en su obra Exploratory Data Analysis (1977).

El EDA supone no sólo un conjunto de nuevas técnicas, sino sobre todo una nueva actitud de los estadísticos ante los datos a estudiar. Antes se adaptaban los datos observados al esquema de unas distribuciones de frecuencia teóricas (como la de Gauss) y con ello se obtenían una serie de medidas para su descripción y su análisis, tales como la media, la varianza, el coeficiente de correlación de Pearson, etc. El EDA partiendo de los propios datos y sin ideas preconcebidas, intenta extraer las estructuras subyacentes en ellos, sin ajustarlos a teorías previas, considerando que "cuanto más conoce un investigador sus datos, más efectivamente los puede emplear para desarrollar, probar y refinar sus teorías". En pocas palabras, se trata de dejar hablar a los datos por sí mismos.

Tal actitud general se puede concretar en varios principios básicos (HARTWING, F. y DEARING, B., 1979):

- Escepticismo ante las medidas estadísticas que intentan resumir una variable, ya que al mismo tiempo que nos muestran alguna de sus características también nos ocultan otros aspectos de ella.
- Apertura ante la posibilidad de encontrar "estructuras" subyacentes en los datos, distintas a las que podíamos presuponer.

El EDA insiste más en la "estructura" que hay en los datos, en contraposición a la insistencia de la Estadística clásica

en la idea de probabilidad. Evidentemente, el concepto de estructura es más intuitivo que el de probabilidad, lo que facilita la enseñanza del EDA frente a la Estadística clásica. Por otra parte, el EDA se aplica a poblaciones completas y no se considera la existencia de muestras, lo que está de acuerdo con una característica usual de muchos problemas geográficos: la dificultad conceptual de definir lo que es población y de lo que sería una muestra extraída de ella.

Varios temas son importantes en el enfoque del EDA:

- Resistencia: La tendencia a elaborar medidas estadísticas que sean insensibles a la existencia de valores anómalos y extremos en los datos. Se presta atención sobre todo al cuerpo principal de los casos, olvidando o minusvalorando la importancia de los casos extremos ("outliers").
- Residuos: El análisis de los datos nunca está completo sin haber examinado cuidadosamente los "residuos", los aspectos no explicados por el análisis realizado, ya que en ellos pueden aparecer "estructuras" no esperadas y útiles para elaborar nuevas teorías.
- Expresividad gráfica: El EDA otorga una importancia central al empleo de técnicas de carácter gráfico o semi-gráfico, que se consideran son más capaces de mostrarnos el comportamiento completo de los datos. En esto el EDA concuerda muy bien con la tradicional insistencia de la Geografía en los procedimientos gráficos y cartográficos.
- Reexpresión: El EDA no privilegia ninguna escala de medida (como se hace habitualmente con la aritmética). Por ello una operación esencial es la reexpresión de los datos cambiando la escala en que se expresan, desde la aritmética usual a la logarítmica u otras, de modo que se simplifique el análisis y se consiga una serie de propiedades deseables como la simetría de su distribución de frecuencias o la linealidad de la relación entre dos variables. La insistencia del EDA en el empleo de distintas es-

calas ha recibido críticas de algunos geógrafos (GOULD, P., 1970), pero debemos considerar que esto es una práctica habitual en las Ciencias experimentales (por ejemplo, la escala del pH es de carácter logarítmico) y, cada vez más, en las Ciencias sociales se encuentran ejemplos en los cuales los hombres reaccionan ante variaciones no proporcionales de los fenómenos que les afectan.

Partiendo de todas estas ideas se han ido desarrollando un gran número de técnicas analíticas, algunas de las cuales vamos a mencionar a continuación. Un balance detallado de sus características y de sus aplicaciones se pueden encontrar en los trabajos siguientes: VELLEMAN, D. y HOAGLIN, D., 1981; TUKEY, J., 1977; COX, N. y JONES, K., 1981; JONES, K., 1982 y BRADSHAW, R., 1983. Un enfoque más matemático se puede hallar en HOAGLIN, D. y otros, 1983.

**a) Descripción y análisis de una variable:**

- Gráfico "rama-con-hojas" (stem-and-leaf) y gráfico en caja (box-plot) (SIBLEY, D., 1984).
- Resúmenes numéricos resistentes, basados en el orden de las observaciones: mediana, cuartiles, etc.

**b) Relación entre variables:**

- Métodos de ajuste de rectas resistentes a la relación entre dos variables (elaborados a partir de las medianas, y por ello resistentes a la presencia de casos anómalos) que son muy fáciles de aplicar.
- Diversos procedimientos gráficos para el estudio de los residuos de una relación (JONES, K., 1983 y WRIGLEY, N., 1984).

**c) Reexpresión de datos:**

- La llamada "familia de potencias", un conjunto de transformaciones de los datos muy versátiles, capaces de facilitar tanto la simetrización de una distribución de frecuencias de una variable, como la linealización de la relación entre dos variables.

**d) Series espaciales y temporales:**

- Métodos para el "suavizado" (eliminación de los casos extremos de la serie) de valores mediante el procedimiento de las medianas móviles.

**e) Tablas de datos:**

- El procedimiento llamado de "sustracción de medianas" permite realizar con gran facilidad un análisis semejante al de Varianza pero conservando la propiedad de la resistencia (ANDERSON, E. ; COX, N., 1978).

**f) Análisis multivariado:**

- Se ha propuesto un procedimiento que permite la elaboración de una ecuación con varias variables explicativas (semejante a una ecuación de regresión múltiple), pero realizada "paso a paso", analizando de continuo los diversos residuos y reexpresando siempre que sea necesario las nuevas variables antes de ser introducidas en la ecuación final (HARTWING, P. y DEARING, B., 1979).
- El gráfico BILOT para matrices multivariantes que facilita la inspección de las características de los datos (GABRIEL, K., 1981).

En resumen, tenemos con el EDA un conjunto variado y coherente de técnicas que, por un lado, pueden sustituir si es necesario a las más clásicas de la Estadística paramétrica, y por otro lado, las pueden complementar cuando sea necesario.

**4. VENTAJAS DIDACTICAS DEL EDA**

La enseñanza del EDA a los estudiantes de Técnicas cuantitativas en Geografía presenta una serie de ventajas de tipo práctico y didáctico. Evidentemente no está aquejado de alguno de los graves problemas que antes hemos señalado que afectan a la Estadística clásica (en especial su subordinación a la distribución de Gauss), pero, más aún, su aprendizaje suele ser más fácil que el de

la Estadística. Se requiere, en general, menos conocimientos previos de matemáticas, sus conceptos son más intuitivos y sus resultados y métodos más próximos a las necesidades habituales de los geógrafos.

En concreto, nuestra experiencia en la enseñanza de estos procedimientos en los últimos dos cursos, creemos que ha sido francamente positiva. Hemos combinado el estudio del EDA con la Estadística clásica descriptiva, y de este modo creemos que los alumnos han conseguido entender más adecuadamente las ventajas e inconvenientes de cada enfoque, sin convertir en un remedio absoluto ninguna medida estadística.

Para comprobar con más exactitud esta impresión personal, hemos llevado a cabo una encuesta, tanto a los alumnos del año 1983-84, como a los del 1984-85. Nuestra intención era, entre otras cosas, comprobar el grado de dificultad atribuido por los estudiantes al estudio de ciertos temas, unas veces (en 1983-84) usando la Estadística clásica, y otras veces (en 1984-85) mediante el EDA. El cuadro I muestra los resultados. Está claro que se ha producido una disminución del grado de dificultad atribuido por los estudiantes que han empleado el EDA. Lo que corrobora, aunque no de un modo definitivo, muestra opinión de la mayor facilidad de estudio del EDA.

## CUADRO I

### NIVEL DE DIFICULTAD SUBJETIVO

TEMAS	ESTUDIADO CON LA ESTADISTICA CLAS.*	ESTUDIADO CON EL EDA*	DIFERENCIA DE MEDIAS
Descripción de una variable	5.16	4.83	0.33
Relación entre dos variables**	6.90	5.96	0.94
Elaboración de un modelo multivariado	8.12	7.74	0.30

\* Media aritmética de todos los alumnos (en torno a 30), medido mediante una escala que varía entre 1 (Mínima dificultad) y 10 (Máxima dificultad subjetiva atribuida por cada alumno a cada tema).

\*\* En este caso la diferencia entre las dos medias es significativa estadísticamente para un nivel de error del 10%.

## 5. CONCLUSIONES

Parece lógico concluir insistiendo en la utilidad del EDA para el análisis de datos geográficos, y sobre todo, en el aprendizaje de los procedimientos adecuados a resolver estas tareas. En cualquier caso, sus posibles inconvenientes (señalados ya por EHRENBURG, A., 1979) serán mejor comprendidos y estudiados con una mayor difusión de estas técnicas.



## BIBLIOGRAFIA

- ANDERSON, E.W. y COX, N.J. (1978): "A comparison of different instruments for measuring soil creep" Catena, 5, 181-93.
- BRADSHAW, R. (1983): "El futuro de la Geografía cuantitativa" Ponencia en el Curso-coloquio sobre Geografía cuantitativa, Oviedo.
- BOSQUE SENDRA, J., RODRIGUEZ RODRIGUEZ, V. y SANTOS, J.M. (1983): "La Geografía cuantitativa en la Universidad y la investigación española" Geocrítica, 44, 47 pp.
- BOSQUE SENDRA, J., CHUVIECO, E. y SANTOS, J.M. (1983 b): "Algunos problemas metodológicos de las técnicas cuantitativas en Geografía humana" Ponencia en el Curso-coloquio sobre Geografía cuantitativa, Oviedo.
- COX, N.J. (1978): "Exploratory data analysis for geographers". Journal of Geography in Higher Education, 2(2), 51-54.
- COX, N.J. y ANDERSON, E.W. (1978): "Teaching geographical data analysis: problems and possible solutions". Journal of Geography in Higher Education, 2(2), 29-37.
- COX, N.J. y ANDERSON, E.W. (1980): "In defence of exploratory data analysis". Journal of Geography in Higher Education, 4(1), 85-89.
- COX, N.J. y JONES, K. (1981): "Exploratory data analysis" en Wrigley, N. y Bennett, R.J. (editores): Quantitative Geography: a British view. Londres, Routledge and Kegan Paul.
- EHRENBERG, A.S.C. (1979): "A note of dissent on data analysis". Journal of Geography in Higher Education, 3(2), 113-116.
- GABRIEL, K.R. (1981): "Biplot display of Multivariate Matrices for inspection of data and diagnosis" en Barnett, V. (editor): Interpreting Multivariate Data, Chichester, J. Wiley.
- GOULD, P.R. (1970): "Is Statistix Inferens the geographical name for a wild goose?" Economic Geography, 46, 439-448.
- HAGGETT, P. y otros (1977): Locational Analysis in Human Geography 2ª Edición. Londres, Arnold.

- HARTWING, F. y DEARING, B.E. (1979): Exploratory Data analysis,. Londres, Sage. 83 pag.
- HOAGLIN, D.C., MOSTELLER, F. y TUKEY, J.W. (editores) (1983): Understanding Robust and exploratory data analysis, Nueva York, J. Wiley.
- JONES, J. (1982): "Using partials residuals to explore relationships" Comunicación al III Coloquio Europeo de Geografía Teórica y Cuantitativa, Augsburg.
- SIBLEY, D. (1984): "A robust analysis of a minority census: the distribution of travelling people in England" Environment and Planning A. 16, 1279-1288.
- TUKEY, J.W. (1977): Exploratory Data Analysis. Nueva York, Addison-Weeley.
- VELLEMAN, D.F. y HOAGLIN, D.C. (1981): Applications, Basics and Computing of Exploratory Data Analysis. Boston, Duxbury Press.
- WRIGLEY, N. (1984): "Quantitative methode: on data and diagnosti" Progress in Human Geography, 567-577.