

## **PROGRAMAS ESTADISTICOS A EXAMEN**

**Vicente RODRIGUEZ RODRIGUEZ**

**Instituto de Economía y Geografía Aplicadas  
Centro Superior de Investigaciones Científicas  
Calle Pinar, 25  
E-28006 Madrid**

### **INTRODUCCION.**

El trinomio Informática-Estadística-Geografía tiene una larga tradición desde que la metodología cuantitativa impuso una forma de entender el análisis geográfico distinto a lo anteriormente instituido. La evolución de la tecnología informática (máquinas y programas) y la abundante información disponible sobre el espacio geográfico han contribuido a su afirmación a lo largo del tiempo. No hay que olvidar tampoco la interdisciplinariedad que se produce en muchas ciencias, y en concreto en las Ciencias Sociales, consistente en el trasvase de métodos e instrumentos de análisis (RODRIGUEZ, 1988, 117-120).

Los métodos estadísticos se configuran como un elemento común al trinomio, de manera que su profución no ha hecho sino aumentar, a veces de forma casi inabarcable. Como ha señalado MAGUIRE (1989, 49) estos métodos tienen un carácter versátil ya que permiten describir, clasificar y sintetizar la información, o establecer inferencias, o, simplemente, "ser más rigurosos en nuestro trabajo" o "dedicar más tiempo a la interpretación que al cálculo de las estadísticas", y ser aplicables a muchos tipos de datos (O'KEEFFE y CLAGGE, 1986, 7). Semejante alteración en el análisis geográfico tiende a reforzar en muchos geógrafos "las furtes tradiciones positivistas" (UNWIN y DAWSON,

1985, 229) o los que es más grave una restricción del "campo del pensamiento de tal modo que sólo se toman en consideración aquellos problemas para los que se necesita la máquina" (GOULD, 1987, 25).

Sin profundizar en cuestiones filosóficas, es una realidad incuestionable la presencia entre los geógrafos y paquetes de programas estadísticos y gráficos que hacen diferente la investigación. Ahora bien, de la abundancia muchas veces nace la inconsistencia. Es decir, la profusión de paquetes no siempre significa que sean adecuados para la investigación en Ciencias Sociales. Esta cuestión tiene cada día mayor importancia como lo demuestra la aparición de algunos análisis de paquetes desde diversos aspectos (SCHRODT, 1984; UNWIN y DAWSON, 1985; SIEGEL, 1985; O'KEEFFE y KLAGGE, 1986; CEBRIAN y BOSQUE, 1987; WOODWARD et al., 1988; MAGUIRE, 1989), además de los efectuados por revistas informáticas (Byte, Pc-World, PcMagazine, PcForum, etc).

El objetivo de este trabajo es intentar dar respuesta a ciertas preguntas que pueden ser consideradas como hipótesis a contestar. Así ¿Tienen los paquetes estadísticos criterios de funcionalidad y de utilidad que los hagan más fáciles de aplicar y más "amigable al usuario"? ¿qué rasgos técnicos dificultan su aplicabilidad?

Para ello requiere disponer de un catálogo suficiente de paquetes y programas, en este caso el libro de WOODWARD et al. Directory of statistical microcomputer software, con 214 productos recogidos por medio de un cuestionario enviado a las empresas productoras. Como fuente tiene a enorme utilidad de permitir analizar los rasgos de una cantidad importante de programas. Tres inconvenientes merman su potencialidad,

a) el referirse casi exclusivamente a productos disponibles en el mercado norteamericano.

b) el ser un catálogo, que no un análisis de los programas y paquetes.

c) estar referidos aproximadamente al año 1986.

A sido necesario deshechar algunos programas que no tenían aplicaciones estadísticas o muy reducida y de dudosa utilidad en Geografía (ej. control de calidad), hasta alcanzar la cifra de 184 productos.

### CRITERIOS DE FUNCIONALIDAD.

En este apartado se presentan los rasgos y las aplicaciones fundamentales de los programas.

Tanto el "tipo de programa" como el "uso" a que se destinan constatan una de las ideas básicas que caracterizan a las aplicaciones informáticas útiles a la investigación en Geografía. Casi la 3/4 partes son "paquetes de análisis estadístico de propósito general" (MAGUIRE, 1989, 53) y con uso mayoritariamente estadístico, frente a la escasa importancia de programas estadísticos específicos (análisis de regresión, de series temporales, de datos categóricos) y a los de análisis específico (sobre todo, control de calidad) (Fig. 1).

En definitiva, son paquetes y programas con una orientación acentuada a la aplicación estadística y escasa hacia la vertiente educativa. Un análisis de las aplicaciones estadísticas incluidas permite apreciar algunos otros hechos notable :

a) la estadística descriptiva y el análisis de correlación y regresión están presentes en el 85% y 70% de todos los paquetes y programas analizados, con unos valores medianos por paquete de 7 y 6 aplicaciones respectivamente (Fig. 2). La misma distribución de los

procedimientos estadísticos, en forma relativamente simétrica, es un indicador de su presencia continua en casi todos los productos.

En el caso de la estadística descriptiva los 7 procedimientos más frecuentes son la media, la varianza y desviación típica, el error standar de la media, la mediana, los percentiles, la asimetría y la curtosis, con otros menos frecuentes como el valor mínimo y máximo, el rango y el coeficiente de variación. Para el análisis de correlación destacan el coeficiente de correlación lineal simple y múltiple, la regresión por pasos y el cálculo y representación de los residuos.

b) en un segundo grupo existen unos valores distintos pero una distribución semejante: así los "procedimientos diversos" en más del 80% de los programas y el análisis de datos categóricos (54%) tienen unos valores medianos de 5 y 4, pero con una distribución en sentido descendente a medida que aumenta el número de procedimientos. Ello significa que el análisis de datos categóricos se centra sobre todo en el análisis de tablas multidimensionales de Ji cuadrado y tests de bondad de ajuste, tratamiento estadístico muy influido por la Sociología y Psicología.

Entre los procedimientos "diversos" destacan la generación de números aleatorios y el cálculo de probabilidades a partir de tests. En esta apartado se deberían encontrar los métodos multivariantes, pero estos están ausentes en los programas, no en los grandes paquetes para microordenadores. Tomando como muestra algunos métodos multivariantes, el análisis factorial sólo está en 15 paquetes 8% del total, el análisis de componentes principales en 12 (6,5%), el análisis "cluster" en 14 (7,6%), o los modelos log-lineales en 6 (3,3%). Únicamente 7 paquetes tienen los tres primeros procedimientos (BMDP, NTSYS, NCSS,

Prodas, SPSS, Statpro y Systat) y en tres, los cuatro métodos (Prodas, SPSS, Systat). Algunos de estos paquetes han incorporado desde 1986, fecha de recogida de información para el catálogo, estos y otros métodos multivariantes.

c) el resto de las aplicaciones estadísticas tienen una implantación más reducida: con la excepción del análisis de varianza, de los tests paramétricos y de las funciones estadísticas (en el 53%, 61% t 49% de los programas respectivamente), el resto tienen escasa representación, tienen un elemento en común, el que su distribución es de tipo bimodal o multimodal (Fig. 3)

En el caso de funciones estadística muchos programas sólo disponen del análisis de la curva normal, mientras otros agrupan 4 funciones clásicas (normal, t, Ji cuadrado y F). Los tests paramétricos esenciales son el análisis de una o varias muestras con iguales o distintas varianzas (4 procedimientos), a los que se suele añadir análisis de medias y sus diferencias, de proporciones y sus diferencias, y de varianzas. Algo semejante sucede con los tests no paramétricos donde el cálculo de los tests de Wilcoxon, Mann-Whitney, Kruskal-Wallis y Spearman son más frecuentes, pero siendo también posible encontrar otros (Kolmogorov, Goodman-Kruskal, Friedman). El análisis de varianza se desdobra entre las varias posibilidades de ofrece ( de una o dos vías), junto con el análisis de covarianza.

Quizás por la importancia de la Econometría, el análisis de series temporales es muy importante, no sólo por el valor mediano sino también por los distintos estratos modales y el amplio rango de procedimientos. Destacan las aplicaciones de medias móviles, de funciones de autocorrelación y de modelos ARMA y ARIMA, con complementos gráficos muy desarrollados. El análisis exploratorio de datos es un conjunto de técnicas con reducida implantación en los

programas del catálogo.

En definitiva, la proliferación de programas con reducido bagaje de aplicaciones estadísticas limita la funcionalidad de la mayor parte de éstos. En cambio los grandes paquetes tienen una oferta mucho más amplia, con una media superior a 10 procedimientos en la mayor parte de las aplicaciones, asociados con un desarrollo gráfico muy notable. También se lalen de esta tónica general los programas estadísticos de propósito específico.

### CRITERIOS DE UTILIDAD.

Se engloban en este apartado aquellos caracteres que favorecen (o no) el uso de programas y paquetes en las gestión y tratamiento de la información y en los productos obtenidos.

Uno de los más importantes tiene que ver con el acceso a las funciones del programa: se conocen dos formas, por medio de menús y de comandos, cada una de ellas con sus ventajas e inconvenientes (SCHRODT, 1984, 67; SIEGEL, 1985, 2). El predominio de programas y paquetes accedidos por menús es claro (Fig. 4) frente a la forma de comandos. La única solución para sortear las desventajas de ambos sistemas es complementarlos en un mismo producto y ello se produce en 9 productos, destacando MicroTSP, Stata/Graphics, Statgraphics y Statpac Gold. Solamente un 10% se conceptúan como programas independientes sin formar parte de un sistema integrado.

Una de las mayores dificultades que puede tener un producto informático es la rigidez en la entrada de información (SCHORODT, 1984, 66). La dependencia de la entrada de datos por el teclado es un hecho evidente para un número de productos que sólo tienen como alternativa la recepción de datos en ASCII, como formato universal (Fig. 4). Ni siguiera otros programas,

muy extendidos en su uso y que organizan y gestionan la información (Lotus y DBase), suplen estas carencias. Realmente un porcentaje amplio de estos programas tienen limitaciones notables para la entrada de datos.

Sin embargo, el tratamiento de los datos está aceptablemente superado ya que sólo en 20% de los programas no permite la manipulación de los datos, previa al análisis, y el 66% está capacitado para operar con datos incompletos ("missing values"). Pero la lentitud en el cálculo estadístico al no tener coprocesador matemático (como sucede en los dos tercios de los programas) puede representar otro elemento negativo. Es cierto que muchas de las aplicaciones no requieren coprocesador, dada la sencillez de los métodos estadísticos.

Otro aspecto necesario es la documentación del programa (manual y ayuda en pantalla). En el primer caso (Fig. 4) los manuales tienen en su conjunto las 5 grandes funciones (tabla de contenidos, índice, tutorial de ejemplos, "sample runs" y fórmulas) que permiten tener informado adecuadamente al usuario en su pretensión de saber qué tiene que hacer en cada momento y cómo tienen que hacerlo. No está, en cambio, tan bien resuelta la disponibilidad de ayuda por pantalla, ya que sólo la mitad de los programas la incorporan reduciendo su versatilidad por ello.

El resultado más evaluable són los gráficos estadístico, tanto en su aspecto de producto resultante del análisis como en su presentación estética (ESCOLANO, 1988, 321). Más de la mitad de los programas realizan gráficos con un cierto grado de complejidad (representación de residuales o de factores), frente a un tercio que sólo hacen histogramas de frecuencia y diagramas x-y, quedando el resto de los productos sin oferta de gráficos complementarios al análisis estadístico. No sólo es esta cierta limitación,

sinó también las operaciones con ficheros (Fig. 4), algo escasas, especialmente las presentaciones en color en el monitor.

Como consecuencia del diseño del catálogo analizado, los programas presentan algunas rigideces (entrada y tratamiento de datos, la no generalización de la ayuda en pantalla, o de coprocesador matemático) que los hacen poco útiles, frente a las estructuras más complejas y versátiles de los paquetes, como han demostrado SIEGEL (1985), O'KEEFFE (1986) o MAGUIRE (1989). Seguramente la capacidad evolutiva de los productos del software habrá eliminado algunos de estos déficits.

### CRITERIOS GENERALES.

Algunos otros parámetros merecen ser considerados a la hora de evaluar la capacidad y aplicabilidad de los programas y paquetes estadísticos.

La capacidad de memoria interna requerida (RAM) demuestra que la mayor parte de los programas sólo necesitan 256 k. (valor modal máximo de 50 productos, el 27%). Junto a los programas con 48, 64 y 128 k. representan el 70% del total del catálogo. Se trata pues, para los datos de 1986, de programas de tamaño muy reducido en el consumo de memoria interna. Sólo los "grandes" paquetes necesitan memorias de 512 y 640 k.

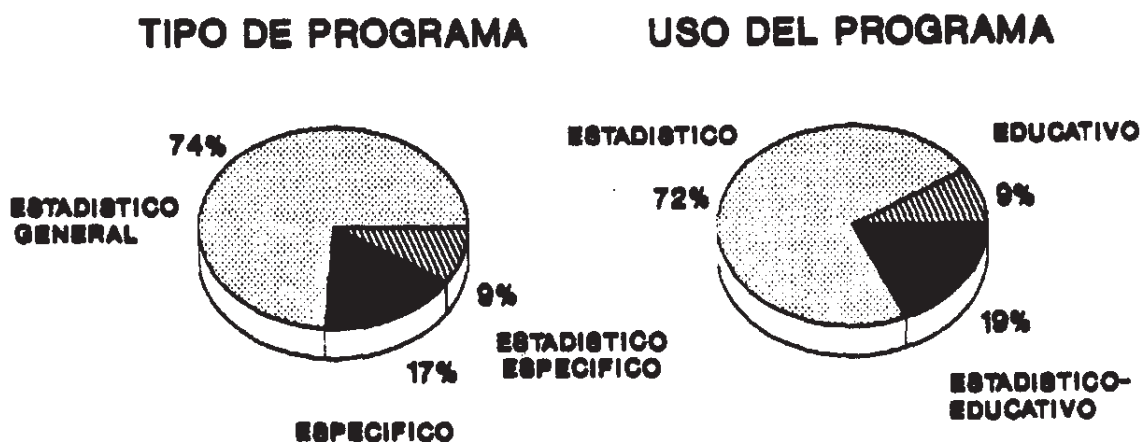
El lenguaje BASIC (37% de los programas) tienen una reducida flexibilidad al no ser un lenguaje avanzado, "no diseñado para programación matemática" (SCHRODT, 1984, 68). Esa condición la cumplen especialmente los escritos en Fortran, Pascal y APL (55 programas, 30%), entre los que se encuentran los paquetes más conocidos y difundidos (BMDP, MiniTab, NTSYS-PC, Prodas, SAS, SPSS, Statgraphics, Systat). Los lenguajes utilizados



diferencian dos grupos de programas, los de aplicación reducida y tratamiento estadístico sencillo (en BASIC), y los de mayor integración y prestaciones.

En cuanto al tipo de ordenador sobre el que funcionan también hay una división entre entornos, aunque más desequilibrada: el de IBM y ordenadores compatibles (67%) y el de APPLE (17%). El resto lo componen un grupo de programas diseñados para ambos tipos de ordenadores. Los paquetes más grandes mantienen esta división ya que sólo Statview y Systat funcionan en entorno APPLE.

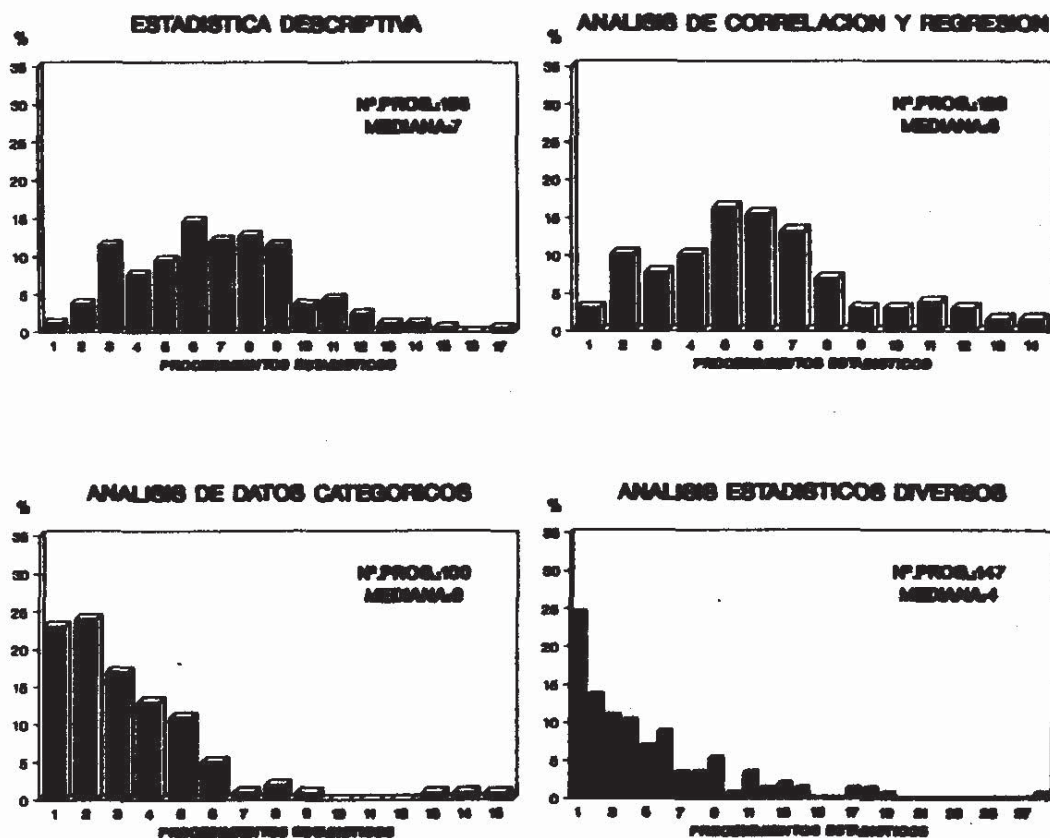
En definitiva, el entorno IBM, por su presencia comercial, determina las aplicaciones estadísticas, aunque muchas de ellas no sean sino sencillas programas para propósitos muy concretos, lejos de una estadística coherente.



FTE.: WOODWARD et al. ELABORACION PROPIA

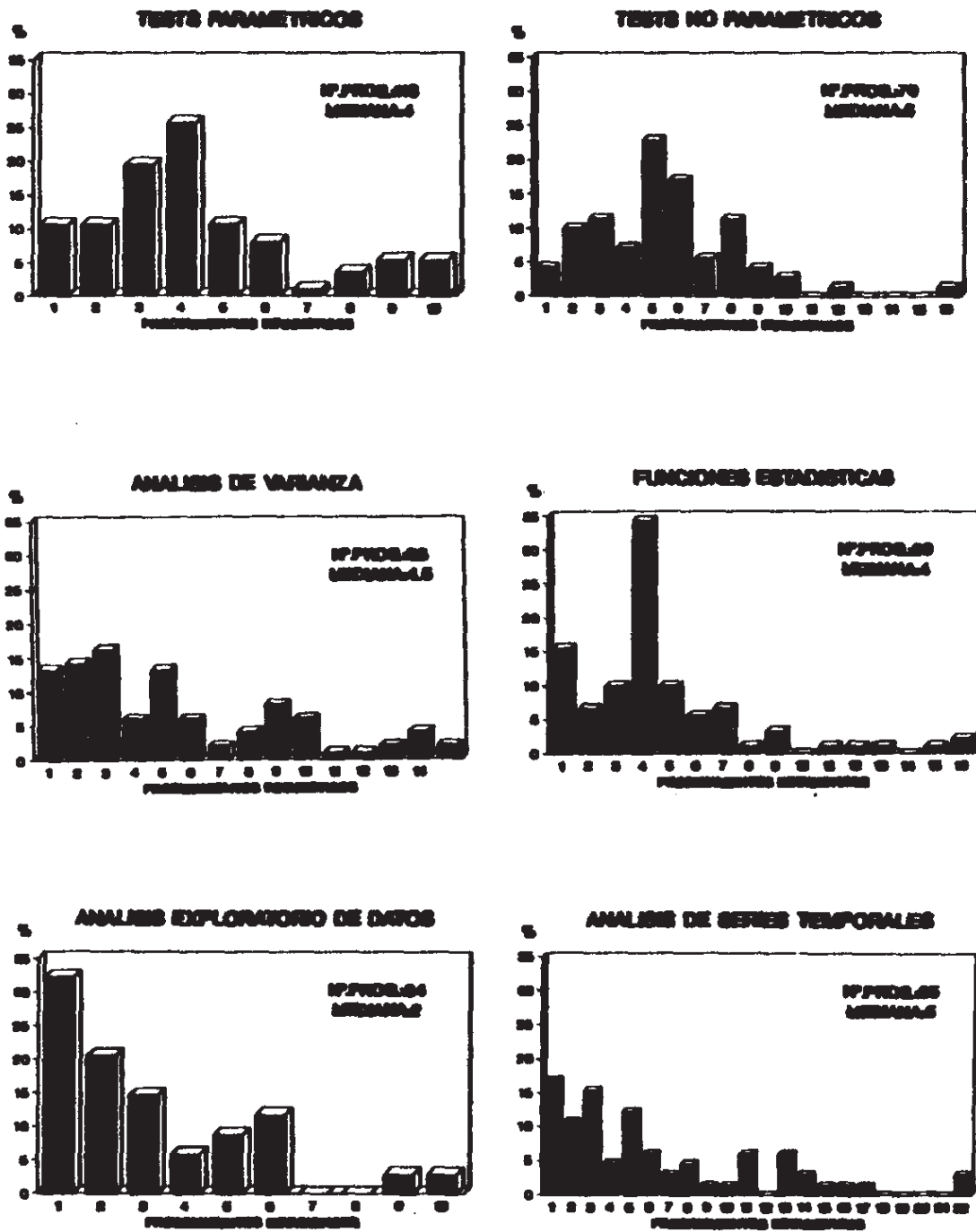
Fig. 1

Fig. 2



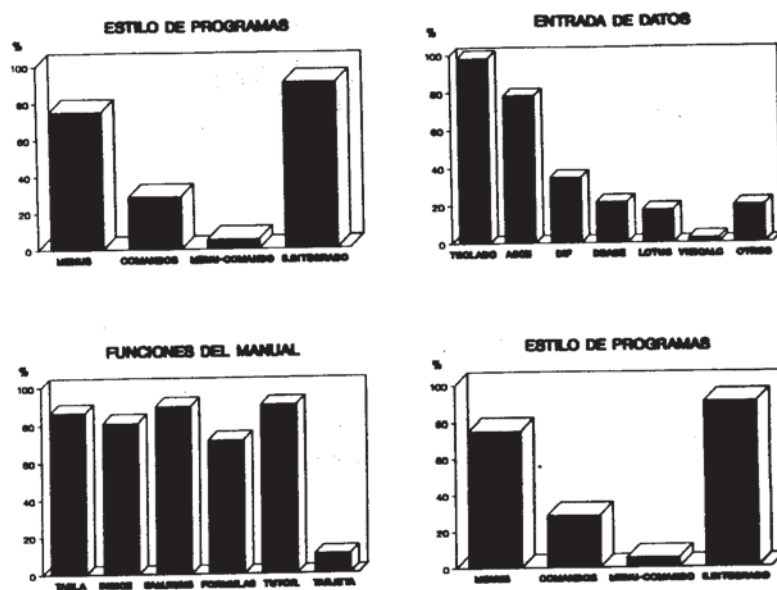
FTE.: WOODWARD et al. ELABORACION PROPIA

Fig. 3



FTE.: WOODWARD et al. ELABORACION PROPIA

Fig. 4



FTE.: WOODWARD et al. ELABORACION PROPIA

## BIBLIOGRAFIA.

BOSQUE, J. et al. (1988): *Aplicaciones de la Informática a la Geografía y las Ciencias Sociales*. Madrid, Síntesis, 319 pgs.

CARPENTER, J; DELORIA, D.; MORGENSTEIN, D. (1984): "Statistica analysis for microcomputers", *Byte*, vol 9, núm. 4, 234-264.

CEBRIAN, J. A. y BOSQUE, J. (1987): "Microordenadores en Geografía". *X Congreso Nacional de Geografía*, AGE, Zaragoza, vol. I, 255-265.

ESCOLANO, S. (1989): "Las representaciones gráficas en Geografía: posibilidades con el ordenador MACINTOSH". *Norba. Geografía VIII-IX. III Coloquio de Geografía Cuantitativa*. Cáceres, Servicio de publicaciones, 317-346.

GOULD, D. J. (1987): "Pensamientos sobre la Geografía". *Geocrítica*, núm. 68, 63 pgs.

MAGUIRE, D. J. (1989): *Computers in Geography*. Harlow, Longman, 248 pgs.

MIDGLEY, H. y WALKER, D. (1985): *Microcomputers in Geography Teaching*. Londres, Hutchinson.

MORENO, A. (1989): "Modelos didácticos y uso del ordenador en los estudios de Geografía", *Norba. Geografía VIII-IX. III Coloquio de Geografía Cuantitativa*. Cáceres, Servicio de publicaciones, 293-315.

O'BRIEN, L. G. (1986): "Statistical Software for microcomputers". *Area*. núm. 18, 39-42.

O'KEEFFE, L. y KLAGGE, J. (1986): *Statistical packages for the IBM/PC family*. New York, MacGraw

Hill, 162 pgs.

OTTENSMAN, J. R. (1985): *BASIC microcomputers programs for urban analysis and planning*. New York. Chapman and Hall.

RODRIGUEZ, V. (1988): "Informática en Geografía y Ciencias Sociales", en BOSQUE, J. et al. (1988): *Aplicaciones de la Informática a la Geografía y las Ciencias Sociales*. Madrid, Síntesis, 319 pgs. (113-124).

SANTOS, J. M. y MUGURUZA, C. (1988): "Microordenadores y análisis estadístico en Geografía", en BOSQUE, J. et al. (1988): *Aplicaciones de la Informática a la Geografía y las Ciencias Sociales*. Madrid, Síntesis, 319 pgs. (141-182)

SCHMID, C. F. (1983): *Statistical graphics: design principles and practices*. New York, Wiley and sons.

SCHRODT, Ph. A. (1984): *Microcomputer methods for social scientists*. Londres, Sage Publications, 96 pgs.

SIEGEL, J. B. (1985): *Statistical software for microcomputers*. New York, North-Holand, 370 pgs.

UNWIN, D. B. y DAWSON, J. A. (1985): *Computer programming for geographers*. Londres, Longman, 252 pgs.

VIDAL, M. J.; BOSQUE, J.; MORENO, A. (1981): "Procedimientos informáticos para el análisis de datos geográficos". *Anales de Geografía de la Universidad Complutense*, núm. 1, 333-344.

WOODWARD, W. A. et al. (1988): *Directory of statistical microcomputer software*. New York, Marcel Dekker, 744.