Modesto ESCOBAR MERCADO
**El análisis de segmentación: técnicas y aplicaciones de
 los árboles de clasificación**

Madrid: Centro de Investigaciones Sociológicas,
 colección Cuadernos Metodológicos, número 39, 239 pp. 2007

Basta con realizar un rápido recorrido por las revistas científicas de Ciencias Sociales para comprobar que las *clásicas* tablas de contingencia van perdiendo presencia y son sustituidas por modelos multivariantes¹. En la explicación de este fenómeno debemos considerar, además de los avances informáticos y la facilidad de acceso a los paquetes estadísticos, la gran difusión de la colección *Cuadernos Metodológicos* del Centro de Investigaciones Sociológicas. Esta colección se ha venido ocupado de *todos* los aspectos relacionados con la investigación social en castellano, del mismo modo que —desde 1976— ha procedido la colección *Quantitative Applications in the Social Sciences* de la editorial Sage en el ámbito anglosajón.

Tras varios libros dedicados a técnicas de análisis de datos de *dependencia*² consideramos que era el momento adecuado para dedicar un trabajo monográfico al análisis de *segmentación*. Mucho más si tenemos en cuenta que en los últimos años se han realizado en nuestro idioma nume-

rosos “acercamientos parciales” a éste mediante artículos de revista y capítulos de libros.

La primera publicación sobre el tema en nuestro idioma fue la traducción de la obra clásica de Mayntz *et al.*, (1976: 282-293), al que siguió el trabajo sobre Sánchez Cuenca (1990) dirigido al ámbito de la investigación comercial. Desde entonces se ha incrementado notablemente el número de investigaciones publicadas que han utilizado el análisis de segmentación (entre otros Rodríguez del Bosque *et al.*, 1998, Riquier *et al.*, 1998: 111-129 y Rufin Moreno 2005) así como los trabajos sobre las características y propiedades de la técnica (entre otros Lamber *et al.*, 1997: 42-55; Luque 2000: 348-382; O'Brien y Durfee, 1994: 36-39; Román González, 2003: 567-230; y Sarabia y Munuera, 1994: 111-124).

El libro objeto de esta recensión supone una actualización de dos trabajos previos del autor, concretamente el *Working Paper* 1992/31 del Instituto Juan March (Escobar 1992) y el artículo “Las aplicaciones del análisis de segmentación: El procedimiento Chaid” publicado en el número 1 de la revista *Empiria* (Escobar, 1998: 13-49). Ahora bien, a nuestro juicio más que una *actualización* supone un *recorrido exhaustivo* por los diferentes tipos de técnicas de segmentación, recogiendo los avances surgidos en los últimos años, todo ello

¹ Obsérvese, por ejemplo, que en los últimos números de la *Revista Española de Investigaciones Sociológicas* publicadas (concretamente el número 123 aparecido en el verano del año 2008) aparecen dos artículos que utilizan la regresión logística.

² Entre otros *Métodos de Análisis Causal* (nº 3), *Regresión Múltiple* (nº 4), *Regresión Logística* (nº 15).

contextualizado dentro de las técnicas de análisis de datos de encuesta.

Pero vayamos con el libro, que es de lo que se trata en este trabajo. El texto, compuesto por nueve capítulos y dos apéndices, comienza con una introducción donde se presentan los principales usos de la segmentación, comparándolos con las utilidades de otras técnicas de análisis multivariantes. Así, entre las páginas 10-11 se señala que el análisis de segmentación “*divide en partes homogéneas una muestra ...detecta qué variables marcan la heterogeneidad y si entre ellas se produce una determinada pauta conjunta en la explicación de la variable dependiente... lleva a cabo una clasificación de los datos... y, por último, es capaz del reconocimiento de pautas*”³. Estas utilidades proporcionan una primera caracterización del análisis de segmentación, que es definido en la página 13 como “una técnica que fragmenta las muestras utilizando un proceso secuencial descendente, que forma grupos homogéneos, exhaustivos y mutuamente excluyentes según los criterios de la variable dependiente (respuesta o criterio), mediante combinaciones jerárquicas de variables independientes (explicativas, clasificadoras o pronosticadores)”.

Dos son, a nuestro juicio, los aspectos más interesantes de la introducción. En primer lugar la “comparación constante” del análisis de segmentación con otras técnicas de análisis multivariable (*conglomerados, Q factorial, Discriminante y Logit*),

lo que supone una contextualización de ésta en la *familia* de las técnicas de análisis de dependencia. Esta comparación permite conocer con precisión las situaciones más adecuadas donde utilizar el análisis de segmentación.

En segundo lugar es destacable el esfuerzo realizado por el autor en describir someramente –y de una forma muy didáctica– los diferentes algoritmos incluidos dentro del análisis de segmentación, presentando las técnicas de clasificación más utilizadas y las principales diferencias entre ellas. Este contexto supone una ocasión excelente para mostrar las principales diferencias entre los cuatro algoritmos sobre los que se profundizará a lo largo del libro: AID, CHAID, CART y QUEST. De modo que tras la lectura de las 16 primeras páginas el lector se encuentra “perfectamente situado” en las *utilidades* de la segmentación, las similitudes y diferencias con otras técnicas de análisis de datos, y las diferencias entre los diferentes algoritmos de segmentación.

A partir de aquí discurren ocho capítulos, que bien podrían dividirse en dos partes. Entre el segundo y el quinto se presentan las diferentes técnicas de segmentación: AID en el segundo capítulo, CHAID y CHAID exhaustivo en el tercero, Árboles de clasificación y regresión–CART en el cuarto, y el algoritmo QUEST en el quinto. En el sexto capítulo se muestran diversas medidas para *evaluar* (y mejorar) el análisis realizado, y en el siguiente se realiza una exposición de las distintas aplicaciones del análisis de segmentación. En los dos últimos capítulos se presentan dos ejemplos con aplicaciones completas del análisis de segmentación.

³ *Pauta* es definida, en este contexto, como un conjunto de características que se diferencia de otro conjunto.

La estructura de los capítulos dos a cinco es similar, al estar dedicados a la presentación de las diferentes técnicas de segmentación⁴. En cada uno se lleva a cabo una exposición en detalle –paso a paso– de las etapas del análisis de segmentación⁵, finalizando con la interpretación de la segmentación resultante.

Es reseñable el gran detalle puesto por el autor en la agrupación (fusión) de las categorías de las variables independientes, y de las pruebas de significación asociadas, mostrando un elevado número de tablas donde se presenta con gran nivel de detalle las diferentes agrupaciones con variables independientes nominales (*free*), ordinales (*float*) o cuantitativas. No menos destacable es el detalle sobre las segmentaciones (“selección de los mejores pronosticadores”), presentando las diferentes variables “candidatas” a segmentar la variable dependiente (o criterio) y las razones por las que se elige una en vez

del resto. Cada uno de estos capítulos prosiguen con los diferentes criterios empleados como “reglas de parada”; bien considerando el número de niveles (filtros de nivel), los tamaños muestrales de cada nodo (filtros de tamaño), o la ausencia de significaciones en la fusión de categorías o en la elección de un predictor adecuado (filtro de significación). Se trata, por así decirlo, de una visión “a cámara lenta” del proceso de segmentación, un “desentrañar” los rápidos procesos realizados por el análisis para localizar la mejor segmentación. Aunque tres párrafos más arriba hemos hablado de una similar estructura del segundo al quinto capítulo, existen ligeras diferencias en el último apartado de cada uno: así el segundo capítulo (AID) finaliza con una explicación de la interacción, el tercero (CHAID) con los indicadores de éxito de la segmentación, el cuarto (CART) con las variables competidoras y los valores perdidos, y el quinto (QUEST) con una comparación entre los cuatro algoritmos. Este último (apartado 5.5, páginas 113 y 115) es de gran importancia al mostrar una síntesis precisa de las analogías y diferencias de los algoritmos explicados anteriormente.

A partir de aquí se produce en el lector una sensación de *relajo* puesto que –explicadas las técnicas de segmentación– cree que el resto del libro se ocupa a “aspectos complementarios”, si bien a medida que transcurre la lectura se aprecia que esta *complementariedad* es menor que la que preveía⁶. El capítulo 6 está dedicado a la

⁴ Por este motivo argumentamos, en el párrafo anterior, que el libro puede dividirse en dos partes; por la similar estructura de los capítulos 2-5.

⁵ Estructura del proceso de segmentación:

Preparación (selección) de variables.

Agrupación de las categorías de las variables independientes (fusión de categorías).

Selección de la variable que realiza la división más heterogénea de la variable dependiente; elección de los mejores pronosticadores.

Agrupación –de nuevo– de las categorías de las variables independientes (dentro de esta división).

Nueva selección de la variable que realice la división más heterogénea.

Finalización del proceso de segmentación:

Filtros de significación.

Filtros de tamaño (muestral).

Filtros de nivel.

⁶ Esto es, es “más fundamental” de lo que era previsible.

evaluación—y mejora—de la segmentación, donde se presentan diversos indicadores que miden la calidad de la segmentación realizada: riesgo, validación y poda de árboles (únicamente disponible en los algoritmos CART y QUEST)

En el séptimo capítulo, que quizás hubiera podido ser el primero, se enumeran las diferentes aplicaciones del análisis de segmentación. Se comienza con las raíces etimológicas del término para proceder a continuación con las diferentes utilidades de la técnica: utilidad descriptiva (con el fin de detectar grupos heterogéneos), utilidad exploratoria (búsqueda de las mejores asociaciones de los predictores con la variable dependiente)⁷, utilidad explicativa (modelos espúreos y modelos multicausales), clasificación de datos (fundamentalmente el algoritmo CART) y predicción. Aunque el análisis de segmentación “fue concebido y es utilizado fundamentalmente con una finalidad exploratoria”, como se señala en la página 148, permite también realizar explicaciones, predicciones y clasificaciones de datos.

Se ha señalado al comienzo del párrafo anterior que el capítulo siete podría haberse colocado al principio del libro porque ahí se presentan —con gran nivel de exhaustividad— las utilidades de la técnica, si bien es verdad que difícilmente pueden explicarse las utilidades de la segmen-

tación si previamente no se conocen sus propiedades y características. Ese motivo explica su ubicación actual, aunque consideramos que los lectores conocedores de la técnica debieran comenzar la lectura por el séptimo capítulo.

En los dos siguientes capítulos se presentan dos ejemplos de investigaciones completas del análisis de segmentación realizadas en nuestro país, concretamente en el ámbito de la afiliación y movilización sindical (capítulo 8) y en el de las identidades múltiples (capítulo 9). Se trata de dos trabajos, ya publicados, que ilustran —de forma aplicada— lo explicado a lo largo del libro. Se incluyen, según argumenta el autor “con objeto de presentar una perspectiva global y aplicada”.

El libro finaliza con dos apéndices. En el primero se presentan las etapas para llevar a cabo el análisis de segmentación, y seguidamente diversos programas de ordenador que realizan segmentaciones, con recomendaciones sobre su manejo y utilización. Un pequeño comentario respecto a las etapas mostradas en este apéndice, al tratarse de una explicación más *general* que las realizadas entre los capítulos dos y cinco, al considerar también la preparación de los datos: propiedades de las variables, elección de variables dependientes e independientes, etc.

En el segundo apéndice se presenta un glosario de los términos utilizados, una idea excelente para las personas que se enfrentan por primera vez a esta técnica. El libro termina con una bibliografía comentada, como es tradicional en esta colección. Destacamos, dentro de ésta, las numerosas referencias de aplicaciones concretas del análisis de segmentación.

⁷ “1) Detectar que variables son útiles para describir adecuadamente las variables dependientes, 2) descubrir qué valores de una variable predictora son homogéneos en relación con la variable a explicar, y 3) poner de manifiesto interacciones entre las variables independientes” (página 141).

Terminaremos con una breve valoración del libro destacando, en primer lugar, la *facilidad* para leerlo. De hecho la primera lectura del libro la realicé en los “tiempos de espera aeroportuarios” en un viaje laboral entre Pamplona y Cádiz, y el texto me “enganchó” de tal modo que a punto estuve de perder uno de los trayectos. La segunda lectura, más reposada y realizada un año más tarde, me lleva a la constatación que se trata de un gran trabajo, y esto es lo que me animó a realizar esta recensión. Hablamos de *Gran trabajo* no por la extensión del mismo, uno de los trabajos más voluminosos de la colección Cuadernos Metodológicos (239 páginas para una colección que suele rondar las 100–150 páginas), sino por la exhaustividad con la que se aborda el tema, la gran precisión de las explicaciones.

Otro de los méritos del trabajo es el amplio especto de público al que se dirige, considerando tanto al lector que se inicia con la técnica como al que tiene alguna experiencia. El texto parte de que el lector carece de conocimientos, pero aborda el tema con tal profundidad que el investigador con conocimientos detiene su lectura en determinados aspectos como la explicación de la interacción, los modelos multicausales, etc.

No menos importante es que todas las explicaciones del libro se basan en datos reales de investigación, lo que supone un gran desafío porque normalmente no suelen obtenerse resultados tan “perfectos” como los datos preparados para la docencia. Además se trata de una información que –en su mayor parte– es fácilmente accesible al lector, lo que permite replicar los análisis realizados. La mayor parte del

libro se ha realizado con investigaciones realizadas por el Centro de Investigaciones Sociológicas, disponibles a todos los investigadores. Así en los capítulos 2, 3, 4, 5 y 6 se utilizó el estudio CIS número 2.511; mientras que en el capítulo 7 se emplean datos procedentes de dos investigaciones del autor, un estudio de CIRES sobre el aborto realizado en 1991, y otro del CIS sobre las elecciones generales del año 2.000 (estudio número 2382).

Otro aspecto a destacar es que el autor anima al investigador en varias ocasiones a *desconfiar* de las elecciones automáticas de la segmentación, instándole a adoptar una actitud crítica de los resultados obtenidos. Recomienda examinar los resultados en sus distintas fases con el fin de valorar el comportamiento de los pronosticadores alternativos; elaborando tablas–resumen con los resultados obtenidos en cada etapa (página 67).

En definitiva, consideramos que se trata de uno de los mejores textos –si no el mejor– sobre el análisis de segmentación, considerando tanto la bibliografía en castellano como en inglés. Se trata de un texto elaborado con sosiego y entrega, que presta una gran atención a cada uno de los aspectos explicados. Una aclaración del autor ayuda a comprender esta situación, cuando señala que “...este libro es fruto de más de dos décadas de aprendizaje y empleo de las técnicas de segmentación...” (p. 16).

Respecto a los inconvenientes del libro tan solo señalo uno; relativo a la dificultad de leer algunos árboles por su reducido tamaño al tratar de “encajados” dentro de una página (páginas 127, 133, 144 y 154). Este aspecto, sin duda producido

por un interés en economizar espacio, genera que el lector deba esforzarse en la lectura y comprensión de los árboles, y consideramos que se trata de un esfuerzo que no debe exigirse al lector. Mucho más cuando se trata de un aspecto que fácilmente podría ser corregido. Por ese motivo consideramos que, en caso de realizar una segunda edición, sería importante aumentar el tamaño de los gráficos incluidos en las citadas páginas.

REFERENCIAS BIBLIOGRÁFICAS

- AEDEMO. 1996. "Segmentación", *Auntes del curso de Análisis Multivariable*. Madrid: entre el 12 de febrero y el 28 de marzo.
- Bagozzi, R. 1993. "The CHAID approach to segmentation modeling". *Handbook of Marketing Research*. Massachusetts: Cambridge.
- Escobar, M. 1992. *El análisis de segmentación: concepto y aplicaciones*. Madrid: Centro de Estudios Avanzados en Ciencias Sociales, Instituto Juan March de Estudios e Investigaciones, Estudio/ Working Paper 1992/31.
- Escobar, M. 1998. "Las aplicaciones del análisis de segmentación: el procedimiento Chaid". *Empiria, revista de metodología de las Ciencias Sociales* 1: 13-49.
- Lambert, A., J. Salazar y J.P. Lévy. 1997. "La segmentación jerárquica y el posicionamiento mediante el uso conjunto del algoritmo CHAID y el análisis de correspondencias". *Esic Market* 97: 42-55.
- Luque, T. 2000. "Segmentación Jerárquica". En T. Luque (coord.), *Técnicas de análisis de datos en investigación de mercados: 347-382*. Madrid: Pirámide.
- Madigson, J. 1992. *SPSS/PC+ CHAID. Versión 5*. Chicago: SPSS inc.
- Mayntz, R. Holm, K.; y Hübner, P. 1975. *Introducción a los métodos de la sociología empírica*. Madrid: Alianza.
- O'Brien, T.V. y P. Durfee. 1994. "Classification Tree Software". *Marketing Research* 6: 36-39.
- Riquier, C., R. Kennedy y B. Sharp. 1998. "Behaviours versus demographics as identifiers of CHAID splits: implications for segment formations". *Journal of Segmentation in Marketing* 2: 111-129.
- Rodríguez del Bosque, I. et al., (1998), "Actitudes hacia la modernización del sector comercial minorista", *Distribución y Consumo*, nº 39, abril/ mayo, pp. 34-43.
- Román González, M.V. y Lévy Mangin, J.P. (2003), "Clasificación y segmentación jerárquica", en J.P. Levy y J. Varela (eds.), *Análisis Multivariable para las Ciencias Sociales*, Madrid, Pearson, pp. 567-630.
- Rufin Moreno, R. 2005. "Patrones de crecimiento comercial y esfuerzo de Marketing en las empresas turísticas españolas". *Investigación y Marketing* 87: 13-16.
- Sánchez Cuenca, J. 1991. "La tipología ". En E. Ortega Martínez (ed.), *Manual de Investigación Comercial*: pp. 524-539. Madrid: Pirámide. (Edición original 1981).
- Sarabia, F. y J.L. Munuera. 1994. "Concepto y usos de la segmentación de mercados: una perspectiva teórica y práctica". *Información Comercial Española* 727: 111-124.
- SPSS. 2001. *AnswerTree 3.0 User's Guide*, Chicago, SPSS inc. Hay versión española titulada *Guía del usuario de AnswerTree 3.0*.

VIDAL DÍAZ DE RADA
Universidad Pública de Navarra