

Una comparación Monte Carlo de tres métodos métricos de segmentación con análisis conjunto

Eduardo Picón *

Universidad de Salamanca (Spain)

En este trabajo se realiza una comparación con datos simulados de tres de los métodos de segmentación con análisis conjunto más populares actualmente y que mejor rendimiento han demostrado en estudios previos. De acuerdo con Vriens, Wedel y Wilms (1996) cabría esperar un rendimiento superior del procedimiento de clases latentes. Sin embargo, los resultados obtenidos muestran un peor rendimiento de este método frente a los basados en una metodología conglomerativa combinada jerárquico + k-means.

Uno de los criterios que ha demostrado ser más relevantes para segmentar mercados de forma efectiva son los beneficios que los consumidores obtienen de los atributos de un producto o servicio (Wedel y Kamakura, 1998). Para evaluar estos beneficios, la herramienta analítica más utilizada actualmente tanto en contextos académicos como aplicados es el análisis conjunto (Green y Srinivasan, 1978; 1990; Varela, 2000; Varela, Rial y García, 2003). El objetivo fundamental de esta técnica consiste en proporcionar información sobre la *estructura de las preferencias* que los sujetos manifiestan por los distintos atributos que configuran un producto. De ahí que el análisis conjunto se utilice frecuentemente para segmentar mercados cuando se desea conocer la estructura de las preferencias que muestran diferentes grupos de sujetos. A partir de sus evaluaciones en una serie de perfiles, el análisis conjunto obtiene unos índices métricos, denominados *utilidades parciales*, que proporcionan una estimación cuantitativa de cuál es el impacto de cada una de las características definitorias del producto en el juicio de preferencia del consumidor (en su decisión de compra). Sumando

* AGRADECIMIENTOS: El autor agradece a Jay Magidson, de la empresa *Statistical Innovations Inc.*, el haberle facilitado una versión de su programa *LatentGOLD*, necesario para realizar la segmentación de clases latentes. Y aprecia sinceramente los pertinentes y enriquecedores comentarios realizados por dos revisores anónimos a una versión anterior de este trabajo. CORRESPONDENCIA: Eduardo Picón Prado. Facultad de Psicología. Universidad de Salamanca. Avda. de la Merced, 109-131. 37005. E-mail: picon@usal.es

las utilidades parciales de cada uno de los niveles que componen un perfil obtendremos la utilidad total o preferencia por ese producto.

La estimación de las utilidades se realiza habitualmente mediante regresión OLS, por lo que, en realidad, sólo necesitamos calcular los coeficientes parciales de regresión β . El problema es que hay ocasiones en las que estimar un único conjunto de β para todos los sujetos puede ser problemático o engañoso. Por ejemplo, la Figura 1 representa un resultado relativamente común en estudios conjuntos: la utilidad estimada para un producto hipotético en función del atributo precio.

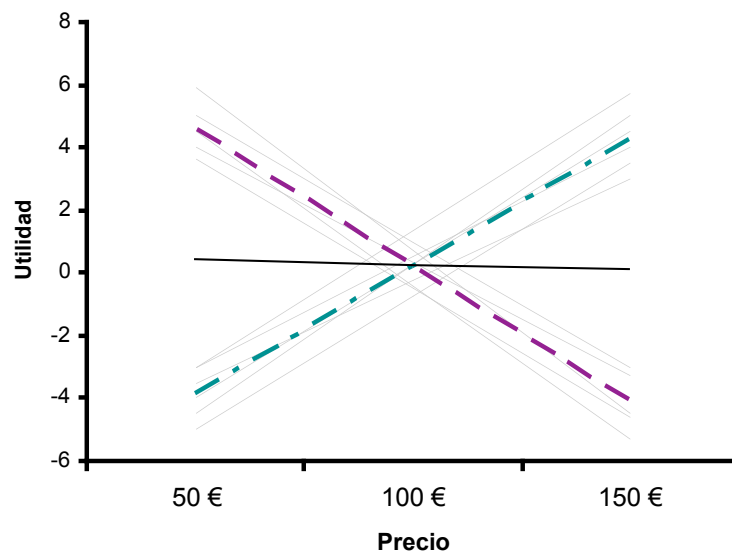


Figura 1. Rectas de regresión para el total de la muestra (línea continua) y segmentando en dos grupos (líneas punteadas) (A partir de DeSarbo, Oliver y Rangaswamy, 1989).

Puede observarse cómo para un grupo de consumidores a mayor precio menor preferencia (en línea con las teorías microeconómicas clásicas), mientras que el otro grupo se comporta de modo inverso, reaccionando positivamente al precio como indicador de calidad. Casos como estos exigen segmentar la muestra y calcular una recta de regresión independiente para cada grupo. Este problema, que podríamos denominar *regresivo-conglomerativo*, se ha intentado resolver de varias maneras.

El enfoque tradicional ha consistido en plantear dos etapas claramente diferenciadas (Picón y Varela, 2000): En una de estas etapas se segmenta la muestra, bien mediante alguna variable de interés seleccionada previamente por el investigador (segmentación *a priori*) o bien mediante una batería de

variables introducidas en un análisis de conglomerados (segmentación *post hoc*). En la otra etapa se aplica un análisis conjunto con el fin de estimar las utilidades. El problema es que estos procedimientos bifásicos optimizan dos criterios diferentes y no necesariamente relacionados entre sí. Por ejemplo, el análisis de conglomerados que se utiliza en el procedimiento *post hoc* trata de minimizar la razón entre la varianza intra e intersegmental, mientras que el análisis conjunto minimiza la suma de los errores al cuadrado.

Intentando solventar este y otros problemas, en los últimos años se han propuesto una serie de métodos alternativos, denominados genéricamente *métodos integrados* de segmentación conjunta, en los que la estimación de las utilidades y la formación de los segmentos se realiza simultáneamente, tratando de optimizar un único criterio de interés (normalmente maximizando una función de verosimilitud logística).

El gran número de procedimientos de segmentación conjunta propuestos, unido a la escasez de estudios comparativos y de resultados concluyentes, ha servido hasta ahora para poco más que confundir al investigador aplicado que busca identificar segmentos de mercado en base a las preferencias de los consumidores. Así, ocurre con frecuencia que un procedimiento es elegido en base a aspectos subjetivos tales como su mayor disponibilidad, su conocimiento casual o incluso las “tendencias” del momento y no tras haber realizado un análisis exhaustivo de sus puntos fuertes y débiles.

Este trabajo es una extensión y profundización del de Picón y Varela (2000) tomando como referencia el artículo de Vriens, Wedel, y Wilms (1996), con el objetivo fundamental de analizar mediante datos simulados tres de los métodos de segmentación conjunta que mejor rendimiento y mayor aceptación están teniendo por parte de los investigadores tanto académicos como aplicados.

En concreto, en este estudio se han comparado tres métodos métricos de segmentación conjunta:

SPHU: la *Segmentación Post hoc con Utilidades*.

SPHP: la *Segmentación Post hoc con Preferencias*.

SCL: la *Segmentación de Clases Latentes* de DeSarbo et al. (1992).

Los primeros dos constituyen los principales métodos clásicos de segmentación con análisis conjunto y son ampliamente conocidos y utilizados en contextos aplicados. Por su parte, el procedimiento de clases latentes es el que está acaparando en la actualidad una mayor atención por parte de los investigadores, ya que algunos estudios (fundamentalmente el de Vriens, Wedel y Wilms, 1996) han encontrado que presenta un rendimiento superior en términos de estimación de parámetros (identificación de segmentos). Además, constituye un buen exponente de los métodos integrados y presenta la enorme ventaja práctica de estar disponible comercialmente desde hace muy poco tiempo en sencillos y asequibles programas de ordenador elaborados en entorno Windows.

DESCRIPCIÓN DE LOS MÉTODOS DE SEGMENTACIÓN

La *Segmentación Post Hoc con Utilidades (SPHU)* conlleva un primer nivel de análisis individual de las preferencias de los consumidores mediante análisis conjunto. En una segunda fase, mediante un análisis de conglomerados, los consumidores se agrupan en segmentos basados en la similaridad de las utilidades obtenidas en el paso anterior. Por lo tanto, en este método la fase de asignación de los sujetos a los grupos y la fase de aplicación del análisis conjunto están necesariamente relacionadas gracias a la aplicación de un algoritmo conglomerativo directamente sobre las utilidades de los sujetos. El análisis de conglomerados permite obtener los mejores segmentos (máxima homogeneidad intrasegmental y heterogeneidad intersegmental) respecto a las utilidades de los encuestados, asegurándonos de que al menos la formación de los segmentos es la óptima. Lo difícil será ahora perfilarlos y conseguir que sean interesantes desde un punto de vista estrictamente estratégico (en el sentido de identificables, accesibles, sustanciales, etc).

Pero este método de segmentación conjunta también presenta limitaciones de tipo metodológico. La primera de ellas es su incapacidad para estimar cualquier modelo que esté hiperparametrizado a nivel individual debido al uso de diseños de bloques (utilizados a veces para reducir la sobrecarga informativa del sujeto). Esto significa que no puede emplearse, por ejemplo, con modelos como el de Green, Carroll, y Goldberg (1981). Una segunda limitación tiene que ver con el uso que hace de dos técnicas analíticas distintas, cada una de las cuales optimiza diferentes criterios. En la etapa conjunta, si por ejemplo se recurre a la estimación de los parámetros mediante Regresión OLS, se intenta minimizar la suma de los errores al cuadrado. Y en la fase conglomerativa se intenta maximizar la razón entre las varianzas entre e intra segmento. Por último, está la limitación de depender de las (en principio) poco fiables estimaciones individuales. Agrupar a los consumidores en base a estas utilidades individuales puede llevar a clasificaciones erróneas de los sujetos a los segmentos. Aunque en estudios de simulación no se ha detectado que esto sea un problema importante (Vriens, Wedel, y Wilms, 1996), si realmente nos preocupa la fiabilidad de las estimaciones podemos optar por dos métodos alternativos: o utilizar enfoques bayesianos para el cálculo de las utilidades individuales o recurrir al siguiente tipo de segmentación.

Green y colaboradores (Green, Carmone, y Wind, 1972; Green y Srinivasan, 1978) proponen un método alternativo denominado genéricamente *Segmentación Post Hoc con Preferencias (SPHP)*. En la primera fase los consumidores se agrupan mediante análisis de conglomerados en función de sus valores de preferencia, mientras que en la segunda fase se estiman distintos modelos conjuntos para cada segmento identificado. Por lo tanto, aquí primero se recurre a un análisis de conglomerados aplicado sobre las preferencias originales de los sujetos y después a un análisis conjunto para obtener las estimaciones de preferencia de los sujetos (Figura 2). La principal ventaja de este procedimiento es que logra mejorar la fiabilidad de las utilidades, ya que se estiman a nivel de cada uno de los segmentos obtenidos

tras el análisis de conglomerados. Sin embargo, mantiene las dos primeras limitaciones metodológicas del método anterior.

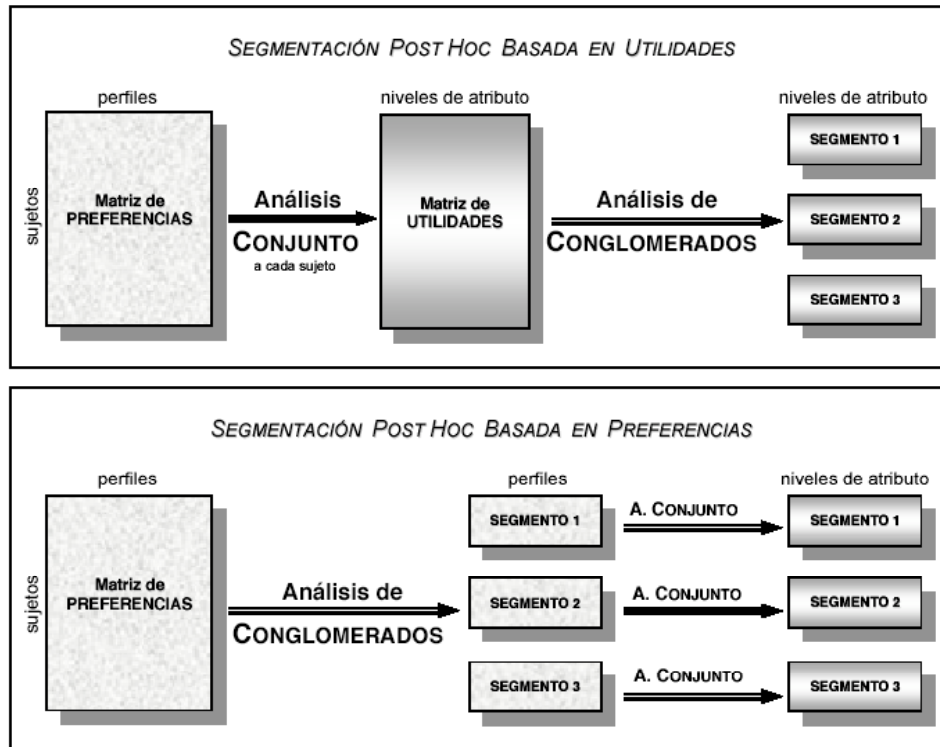


Figura 2. Diferencias entre la Segmentación Post Hoc Basada en las Utilidades y la Basada en las Preferencias.

A la hora de aplicar los análisis de conglomerados en los dos procedimientos *post hoc* de este estudio, se ha seguido la metodología conglomerativa en dos fases propuesta originariamente por Punj y Stewart (1983; véase también Ketchen y Shook, 1996, o Picón, Varela, y Real, 2003) con el fin de aprovechar al máximo los puntos fuertes de los diferentes métodos conglomerativos descritos hasta la fecha. Tras realizar una selección cuidadosa de las variables a incluir en el análisis, se recurre en un primer momento a uno de los métodos jerárquicos que ha demostrado mejor rendimiento (Ward o vinculación intergrupos) con el fin de obtener una primera aproximación a la solución final. Examinando los resultados de este análisis preliminar, el investigador puede determinar tanto el número de conglomerados adecuado (analizando visualmente el dendrograma) como sus centros iniciales (que serán las medias de cada uno de los conglomerados obtenidos en esta fase). Con esta información se ejecuta un nuevo Análisis de Conglomerados, pero esta vez mediante un procedimiento iterativo como el k-medias, para obtener la solución final. Autores como los propios Punj y

Stewart (1983) han encontrado que los procedimientos iterativos son más robustos que cualquiera de los jerárquicos, pero siempre que se especifiquen previamente tanto el número de conglomerados como sus centros iniciales (por ejemplo, mediante un procedimiento jerárquico).

La *Segmentación de Clases Latentes (SCL)*, también conocida como *regresión de mezclas finitas*, ha sido propuesta para el análisis conjunto métrico por DeSarbo, Wedel, Vriens, y Ramaswamy (1992) como una generalización de la metodología de clase latente de DeSarbo y Cron (1988). Este método parte del supuesto de que las preferencias de los sujetos conforman una población que es, de hecho, una mezcla de diversos segmentos en proporciones desconocidas, razón por la cual se desconoce *a priori* el segmento al que pertenece un sujeto concreto. La segmentación de clases latentes asume que la suma de todas las proporciones es 1, y que cada uno de los segmentos se distribuye según una función perteneciente a la familia exponencial multivariada. En ejemplo de mezclas finitas aparece en la Figura 3, en la que se muestra una función de densidad observada (en negro) y sus dos funciones normales subyacentes.

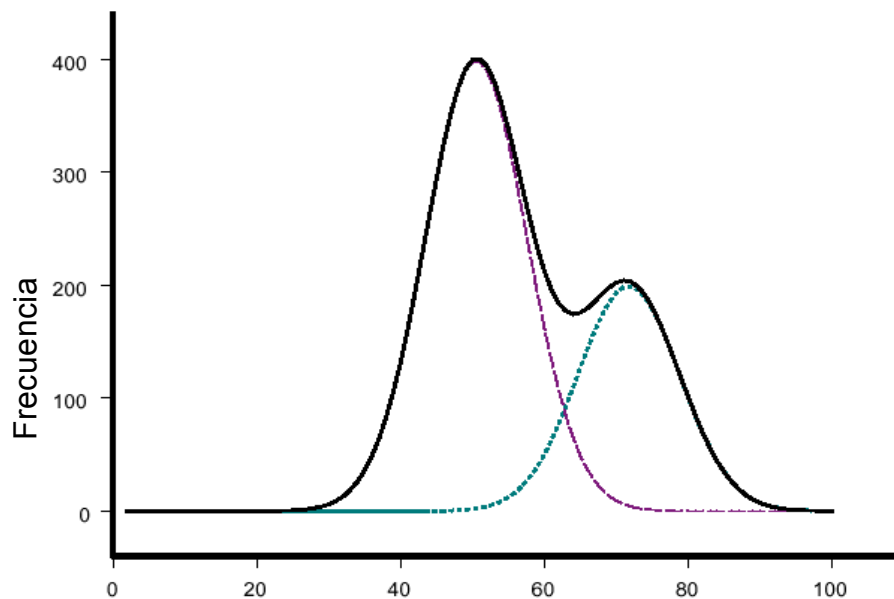


Figura 3. Ejemplo esquemático de una función de densidad observada y sus dos funciones subyacentes.

El propósito es separar las muestras, identificando el número de segmentos y estimando los parámetros que definen a cada uno de ellos, que no son otros que un vector de coeficientes de regresión, y un parámetro de dispersión. La estimación de los parámetros se realiza tratando de maximizar una función log-verosímil mediante el algoritmo de Expectación-Maximización (Dempster, Laird, y Rubin, 1977).

El modelo puede describirse de forma resumida del siguiente modo. Sean:

- n = los consumidores, $1, \dots, N$
- i = los perfiles, $1, \dots, I$
- j = el conjunto de características o atributos, $1, \dots, J$
- s = los segmentos, $1, \dots, S$
- x_{ij} = la variable *dummy* del diseño conjunto para el j -ésimo atributo del i -ésimo perfil. $x = 1, \dots, X$
- $y_{in} = r_{in}$ = la respuesta de preferencia del n -ésimo sujeto al i -ésimo perfil
- β_{js} = el parámetro de regresión para el j -ésimo atributo del segmento s .

Supóngase que en un estudio con análisis conjunto se han recogido y datos de preferencia de N sujetos respecto a I perfiles. Entonces, para cada segmento s , las respuestas de preferencia pueden expresarse como:

$$y_{nis} = \beta_0 + \sum_{j=1}^J \beta_{js} x_{ijs} + e_{nis} \quad (1)$$

Se asume que en cada segmento los sujetos presentan idénticas utilidades parciales para los niveles de atributo, con los errores $e = N(0, \sigma^2)$ idéntica e independiente distribuidos. Las β_s son los coeficientes de regresión (utilidades parciales) específicos del segmento s y los π_s son las proporciones desconocidas en que se reparten los segmentos subyacentes. El propósito es estimar $\beta = (\beta_s, \pi_s, \sigma_s)$. Si asumimos que las y_n se distribuyen normal e independientemente, la evaluación de preferencia del sujeto n por cada perfil, condicionado a que n pertenezca al segmento s (esto es, la distribución condicional de y_{ni}), puede conocerse a partir de:

$$\pi_s(y_{ni} | \beta_s, \pi_s, \sigma_s) = \prod_{j=1}^J (2\pi\sigma_s^2)^{-1} \exp \frac{\pi (y_{ni} - \sum_{j=1}^J \beta_{js} x_{ijs})^2}{2\sigma_s^2} \quad (2)$$

Cada distribución condicional y_{ni} se pondera por la proporción de cada clase latente y la suma de todos estos productos proporciona la distribución incondicional (observada) de las preferencias y_i .

$$\pi(y_i | \beta) = \sum_{s=1}^S \pi_s \pi_s(y_{ni} | \beta_s, \pi_s, \sigma_s) \quad (3)$$

Calculando las derivadas parciales respecto a los parámetros y ajustando éstos a cero se obtienen expresiones cerradas para los β_s y γ_s . Por último, para clasificar a los sujetos en cada segmento se calcula la probabilidad posterior de que el sujeto n pertenezca al grupo s mediante la regla de Bayes:

$$P_{ns} = \frac{\beta_s f_s(y_{ni} | \beta_s, \gamma_s)}{\sum_{s=1}^S \beta_s f_s(y_{ni} | \beta_s, \gamma_s)} \quad (4)$$

El método estándar de clasificación se denomina *modal (modal allocation)* y consiste en asignar cada sujeto al segmento que presente las probabilidades más altas. Aunque algunos autores (Vriens, Wedel, y Wilms, 1996) consideran más adecuada a la clasificación solapada (probabilística), en esta investigación se ha recurrido a la clasificación modal fundamentalmente por dos motivos: (1) porque el análisis previo de los datos ha permitido comprobar que prácticamente todos los casos presentan de hecho clasificaciones no solapadas (el 72,83% de ellos puede asignarse a uno de los segmentos con una probabilidad de acertar $p = 1,00$, y el 97,83% con una $p > 0,90$), y (2) porque además así se facilitan los cálculos y la comparación con los otros dos procedimientos *post hoc* que son no solapados.

DISEÑO DEL ESTUDIO MONTECARLO

Aunque los primeros análisis *tipo* Monte Carlo de un procedimiento estadístico datan de finales del siglo XIX (Stigler, 1978), el término fue acuñado oficialmente por físicos del centro atómico de Los Álamos durante la Segunda Guerra Mundial al observar la similitud entre las extracciones que hacían de números aleatorios y algunos juegos de azar (Binder y Stauffer, 1984; Landau y Binder, 2000). Con la generalización de los ordenadores, el uso de datos simulados ha experimentado un gran desarrollo, constituyendo en la actualidad una importante herramienta para el desarrollo de teorías y métodos matemáticos. Hoy en día, muchos investigadores conciben una *simulación o estudio Monte Carlo* básicamente como un experimento que recurre al uso de muestras generadas de forma aleatoria o pseudoaleatoria con el fin de evaluar una expresión matemática (Gentle, 1998; Lewis y Orav, 1989; Rubinstein, 1981). Se repiten los mismos cálculos básicos un número determinado de veces con diferentes datos simulados y luego se efectúan algunos análisis estadísticos sobre el conjunto de las muestras. Los datos de entrada de los diferentes “ensayos” se seleccionan automáticamente mediante un generador de números pseudo-aleatorios, y lo único que necesita ser preespecificado por el investigador es la función de densidad de probabilidad que describe a la muestra.

En estadística, uno de los usos clásicos de los estudios Monte Carlo tiene por objeto obtener distribuciones muestrales de algún estadístico difíciles de lograr analíticamente y evaluar su comportamiento en un número elevado de muestras aleatorias (Mooney, 1997; Pardo y San Martín, 1998; Umesh y Mishra, 1990). Pero en este trabajo utilizaremos esta metodología con otra

finalidad tal vez menos conocida: comparar experimentalmente las propiedades y/o el rendimiento de diferentes técnicas o modelos estadísticos en liza, tal y como hacen, por ejemplo, DeSarbo, Oliver y Rangaswamy (1989), Hagerty (1985), Vriens, Wedel y Wilms (1996), Wedel y Kistemaker (1989), Wedel y Steenkamp (1991) o Wittink y Cattin (1981).

Basándonos en estos últimos estudios, se han seleccionado cinco factores con posibilidad de tener efectos sobre el rendimiento de los tres procedimientos de segmentación comparados en la presente investigación. Cada factor consta de dos niveles. Los factores recogen aspectos clave como los grados de libertad, el tipo de segmentación o el error de los datos, y aparecen recogidos en la Tabla 1.

Tabla 1. Factores incluidos en el estudio Monte Carlo.

factor	niveles
A. Número de perfiles	9 (0); 27 (1)
B. Número de segmentos	4 (0); 8 (1)
C. Porcentaje de varianza de error en las preferencias	5% (0); 35% (1)
D. Homogeneidad de los segmentos	$\sigma^2 = ,05$ (0); $\sigma^2 = ,10$ (1)
E. Similitud entre segmentos	Similares (0); no similares (1)

a. Entre paréntesis aparece la codificación de indicadores (*dummy*).

La combinación de estos cinco factores en un diseño factorial completo da como resultado 32 condiciones experimentales diferentes (2⁵) que posibilitan la estimación de los efectos principales y de cualquier posible interacción. A continuación se han generado un total de 600 casos repartidos aleatoriamente en 32 muestras (una por condición experimental) de entre 17 y 21 sujetos cada una, fijando el número de sujetos de cada nivel en 300 (n/2).

Cuando se trata de comparar el comportamiento relativo de los diferentes métodos de segmentación, los análisis con datos reales presentan una limitación fundamental: la utilización de una única muestra de sujetos. Esto hace que los resultados estén fuertemente influidos por las características particulares del conjunto de datos utilizado, y dificulta o impide el análisis bajo otras circunstancias diferentes a las que fija la propia muestra. Así, puede ocurrir que un conjunto de datos, aún cumpliendo los requisitos de representatividad, no sirva para conocer cómo se comportarían los métodos si se disminuyese el porcentaje de varianza de error de las preferencias o si los segmentos fuesen más homogéneos. Estos y otros aspectos vienen ya determinados por las propias características de la muestra utilizada. Pues bien,

uno de los grandes atractivos de una comparación Monte Carlo es que permite construir “a la carta” tantas muestras aleatorias como necesitemos para analizar en profundidad la influencia de cualquier predictor sobre el rendimiento de los métodos.

De hecho, una de las decisiones trascendentales que se han de tomar cuando se realiza un estudio Monte Carlo consiste en determinar el número de sujetos y de muestras a incluir en el mismo. Algunos análisis utilizan un único conjunto de datos simulados por condición experimental mientras que otros promedian miles de muestras (Gentle, 1998). Y aunque la recomendación generalizada es utilizar “muchos” conjuntos de datos (Mooney, 1997, p. 58), aspecto éste que resulta clave cuando se trata de analizar la distribución muestral de algún estadístico, en el caso concreto de las *comparaciones Monte Carlo* se suele prestar también atención a aspectos como la eficiencia del propio diseño experimental (Gentle, 1998).

Este es el caso de la presente investigación. El programa estadístico empleado para efectuar la Segmentación de Clases Latentes, *LatentGOLD 2.0*, requiere tener un registro (un caso) de cada una de las valoraciones que ha hecho el sujeto de los perfiles. En consecuencia, las descripciones de cada uno de los 9 o 27 perfiles, junto a la calificación que le otorga cada sujeto, han de aparecer en la matriz de datos en filas sucesivas, haciendo que el número final de filas del fichero (y en consecuencia, la carga computacional) se incremente considerablemente. En concreto, para evaluar este modelo con un $n = 600$ se ha tenido que manejar un fichero de $(300 \times 9 = 2700) + (300 \times 27 = 8100) = 10900$ filas.

Muchos otros trabajos similares a éste emplean la denominación *Monte Carlo* con un número muy reducido no sólo de conjuntos de datos sino también de replicaciones¹. Por ejemplo, Andrews, Ansari y Currim (2002) trabajan con 128 condiciones experimentales y una sola replicación por condición experimental ($n = 150$). Vriens, Wedel y Wilms (1996) utilizan un diseño factorial completo que genera 64 conjuntos de datos (32 muestras con $n_1 = 100$ y 32 con $n_2 = 200$). Wittink y Cattin (1981) trabajan con 18 combinaciones factoriales y también con un solo conjunto de datos por condición (cada uno de ellos de tamaño $n = 50$; $n_{\text{total}} = 900$). Wedel y Steenkamp (1991) generan un diseño factorial fraccionado compuesto por ocho condiciones experimentales y un solo conjunto de datos por condición (cuatro muestras de $n_1 = 50$ y otras cuatro de $n_2 = 100$; $n_{\text{total}} = 600$). Y Hagerty (1985), en un trabajo fundamental para entender los avances posteriores en segmentación conjunta, utiliza tan sólo dos muestras hipotéticas (una por cada una de las dos condiciones contempladas) de 120 casos Monte Carlo cada una ($n_{\text{total}} = 240$).

¹ Sin embargo, otros autores consideran erróneo utilizar la etiqueta *Monte Carlo* en estudios con un número tan bajo de muestras simuladas por condición experimental.

GENERACIÓN DE LOS DATOS

Para generar los datos simulados se ha comenzado creando los diseños ortogonales que permitan obtener los 9 y 27 perfiles (Factor A de la Tabla 1). Cualquier muestra que evalúe 9 perfiles se puede obtener a partir de un diseño completamente ortogonal de cuatro atributos con tres niveles (como el que aparece en la Figura 4). Del mismo modo, una muestra que evalúe 27 perfiles se puede generar con un diseño ortogonal de ocho atributos y tres niveles. A cada diseño se le han añadido ocho perfiles para poder validar posteriormente el análisis.

	a	b	c	d	status_	card_
1	1	1	1	1	0	1
2	3	1	2	3	0	2
3	2	2	1	3	0	3
4	1	2	2	2	0	4
5	1	3	3	3	0	5
6	3	3	1	2	0	6
7	3	2	3	1	0	7
8	2	1	3	2	0	8
9	2	3	2	1	0	9
10	1	1	3	1	1	10
11	1	2	3	1	1	11
12	2	1	3	1	1	12
13	1	1	1	3	1	13
14	3	3	2	2	1	14
15	3	1	3	1	1	15
16	2	2	3	1	1	16
17	2	3	1	2	1	17

Figura 4. Diseño fraccionado completamente ortogonal de 9 perfiles (+8 holdout).

Asumiendo que cada atributo está definido por niveles discretos, únicamente será necesario estimar $4(3-1) = 8$ parámetros en el primer caso y $8(3-1) = 16$ parámetros en el segundo caso. Siguiendo la misma nomenclatura que definimos al hablar de la Segmentación de Clases Latentes, para cada uno de estos parámetros y para cada segmento (Factor B de la Tabla 1) se ha procedido entonces a extraer una serie de utilidades verdaderas “raíz”, (β_{xs}), similares para todos los integrantes de un mismo segmento. Los valores han sido generados aleatoriamente con el SPSS (versión 10.0) a partir de una distribución uniforme con rango $\pm 1,7$, siguiendo el trabajo de Vriens, Wedel y Wilms (1996). La Tabla 2 muestra dichos valores.

En el siguiente paso se han añadido ciertas cantidades (β_{nx}) a estas utilidades “raíz” ($\beta_{nxs} = \beta_{xs} + \beta_{nx}$) con el fin de simular la heterogeneidad intrasegmental. Para obtener las β_{nx} , se han extraído $N \times X$ números aleatorios de una distribución normal estandarizada, multiplicándose a continuación por

la raíz cuadrada de ,05 (segmentos homogéneos) o de ,10 (segmentos difusos). Esto ha dado como resultado una serie de utilidades individuales con una varianza intrasegmental de ,05 ó de ,10. Con estas utilidades simuladas, contenidas en el vector \mathbf{u} ($X \times 1$), y los perfiles contenidos en la matriz \mathbf{X} ($I \times X$) hemos generado las preferencias verdaderas para cada sujeto n , de acuerdo con la expresión $\mathbf{U}_n = \mathbf{X}\mathbf{u}_n$. Subsecuentemente, se ha añadido error a esas utilidades verdaderas \mathbf{U}_n (resultando $\mathbf{r}_n = \mathbf{U}_n + \mathbf{e}_n$). Los porcentajes de varianza que representan las condiciones de baja y alta tasa de error se han fijado en el 5% y el 35% respectivamente. Para obtener la varianza del término de error se ha utilizado la siguiente fórmula (Srinivasan, 1975; Wittink y Cattin, 1981):

$$PVE = \frac{\sigma_e^2}{\sigma_e^2 + \sigma_u^2} \quad \text{c} \quad \sigma_e^2 = \frac{PVE}{1 - PVE} \sigma_u^2 \quad (5)$$

donde PVE es el porcentaje de varianza de error, σ_e^2 es la varianza del término de error y σ_u^2 es la varianza de U . Se han generado $N \times I$ números aleatorios a partir de una distribución normal con media cero y varianza σ_e^2 , añadiéndose a las preferencias verdaderas.

Tabla 2. Valores Raíz Utilizados para Calcular las Utilidades Verdaderas Simuladas^a.

segmento	niveles de atributo ^b															
	a1	a2	b1	b2	c1	c2	d1	d2	e1	e2	f1	f2	g1	g2	h1	h2
1	<u>.90</u>	<u>-1,27</u>	<u>.20</u>	<u>1,64</u>	<u>.59</u>	<u>-.12</u>	<u>.03</u>	<u>-.74</u>	<u>-.92</u>	<u>.77</u>	<u>1,30</u>	<u>-1,08</u>	<u>-.75</u>	<u>.66</u>	<u>1,37</u>	<u>-1,28</u>
2	<u>-.52</u>	<u>1,43</u>	<u>-.31</u>	<u>-1,37</u>	<u>-.56</u>	<u>.41</u>	<u>.48</u>	<u>.86</u>	<u>.03</u>	<u>-1,30</u>	<u>-1,44</u>	<u>.06</u>	<u>.78</u>	<u>-.49</u>	<u>.49</u>	<u>1,20</u>
3	<u>-1,00</u>	<u>-1,21</u>	<u>-.18</u>	<u>-.45</u>	<u>-.05</u>	<u>-1,36</u>	<u>.77</u>	<u>-1,57</u>	<u>-1,48</u>	<u>-.32</u>	<u>.23</u>	<u>1,17</u>	<u>-1,07</u>	<u>-.88</u>	<u>-1,06</u>	<u>1,67</u>
4	<u>1,11</u>	<u>.72</u>	<u>.57</u>	<u>.89</u>	<u>1,03</u>	<u>.69</u>	<u>-1,54</u>	<u>-.36</u>	<u>-.77</u>	<u>.56</u>	<u>.58</u>	<u>.34</u>	<u>1,46</u>	<u>1,14</u>	<u>1,52</u>	<u>.13</u>
5	<u>-1,23</u>	<u>-.23</u>	<u>.38</u>	<u>-.71</u>	<u>-1,17</u>	<u>.68</u>	<u>-.52</u>	<u>-.18</u>	<u>1,25</u>	<u>-.51</u>	<u>-.01</u>	<u>.69</u>	<u>-1,00</u>	<u>.93</u>	<u>-1,47</u>	<u>.09</u>
6	<u>-1,52</u>	<u>-1,35</u>	<u>-1,22</u>	<u>-1,55</u>	<u>.41</u>	<u>-1,18</u>	<u>.73</u>	<u>1,46</u>	<u>.55</u>	<u>1,48</u>	<u>-.63</u>	<u>.08</u>	<u>.90</u>	<u>-1,27</u>	<u>.20</u>	<u>1,64</u>
7	<u>.27</u>	<u>-.81</u>	<u>.76</u>	<u>-1,57</u>	<u>-1,36</u>	<u>.79</u>	<u>-.93</u>	<u>-.95</u>	<u>-.59</u>	<u>.12</u>	<u>.03</u>	<u>-.74</u>	<u>-.52</u>	<u>1,43</u>	<u>-.31</u>	<u>-1,37</u>
8	<u>.37</u>	<u>1,54</u>	<u>-.89</u>	<u>-.55</u>	<u>-1,52</u>	<u>.81</u>	<u>-1,29</u>	<u>-1,50</u>	<u>-.56</u>	<u>.41</u>	<u>.48</u>	<u>.86</u>	<u>-1,00</u>	<u>-1,21</u>	<u>-.18</u>	<u>-.45</u>

a. Para la condición 27 perfiles se han utilizado todos los valores de la tabla y para la condición 9 perfiles sólo los valores subrayados. Del mismo modo, para la condición 8 segmentos se han empleado todos los valores mientras que para la condición 4 segmentos únicamente los valores en cursiva. Cada cuadrante de datos ha sido generado independientemente del resto.

b. Por ejemplo, a1 y a2 son los niveles 1 y 2 del atributo a. El nivel 3 se calcula a partir de estos dos niveles (la estimación de las utilidades parciales se realiza mediante regresión con variables ficticias).

FACTORES INCLUIDOS EN EL ESTUDIO

Número de perfiles. El número de perfiles (9 y 27) se ha escogido para que cubra un amplio rango a ambos lados del valor promedio encontrado en aplicaciones conjuntas ($Md = 16$; Wittink y Cattin, 1989), pero sin llegar a los 30 perfiles que constituyen el máximo aconsejable (Green y Srinivasan, 1978). Este factor nos va a permitir evaluar empíricamente el efecto sobre la estimación de parámetros y la predicción de incrementar el número de grados de libertad: Se espera mejores soluciones conglomerativas a medida que aumenta el número de grados de libertad de los diseños. Para hacer variar esta característica, algunos autores optan por modificar el número de perfiles manteniendo al mismo tiempo el mismo número de atributos y niveles (Andrews, Ansari y Currim, 2002; Vriens, Wedel y Wilms, 1996). El problema es que esta estrategia genera diseños que en ocasiones se desvían seriamente de la ortogonalidad y, en consecuencia, con una escasa validez externa. En este trabajo se ha optado por una estrategia diferente: variar el número de atributos del diseño (4 y 8) lo que también permite obtener diseños con diferentes perfiles (9 y 27 respectivamente), que además preservan la ortogonalidad, tal y como hacen la inmensa mayoría de las aplicaciones conjuntas. Aunque un diseño ortogonal de pocos perfiles no genera necesariamente menos grados de libertad que otro más grande (Tabla 3), los dos diseños concretos utilizados en esta comparación sí que maximizan esta característica.

Tabla 3. Grados de Libertad Para Diferentes Diseños Ortogonales (Atributos Discretos con Tres Niveles) Obtenidos con el SPSS.

atributos	niveles	perfiles del diseño ortogonal resultante	grados de libertad
2	3	9	5
3	3	9	3
4	3	9	1
5	3	16	6
6	3	18	6
7	3	18	4
8	3	27	11

Número de segmentos. Se ha incluido aquí una condición de cuatro segmentos y otra de ocho segmentos. Este factor es uno de los que aparecen sistemáticamente en los diferentes estudios teóricos de segmentación conjunta, aunque incluyendo un menor número de segmentos, comúnmente 2 y 4 (DeSarbo, Oliver y Rangaswamy, 1989; Hagerty, 1985; Vriens, Wedel y Wilms, 1996; Wedel y Steenkamp, 1989; 1991). No obstante, en la práctica

resulta extremadamente raro encontrar soluciones conglomerativas compuestas únicamente por dos segmentos, por lo que se ha decidido permutar esta condición por otra mucho más común, que contemple ocho segmentos. Sin duda, un número de segmentos más cercano a las soluciones conglomerativas obtenidas en otros trabajos previos (Picón y Varela, 2000).

Varianza de error. Los porcentajes de error utilizados (5% y 35%) han sido empleados en estudios previos (Andrews, Ansari y Currim, 2002; Wedel y Steenkamp, 1989; Wittink y Cattin, 1981), observándose que cuanto mayor es la varianza peor es el rendimiento de los algoritmos. Vriens, Wedel y Wilms (1996) han encontrado que este factor, relacionado con el grado de compromiso de los sujetos con la tarea conjunta, es el que tiene una mayor influencia en la bondad del modelo, su capacidad predictiva, y su facultad para realizar una buena estimación de las verdaderas utilidades y pertenencias a cada segmento.

Homogeneidad de los segmentos. Las dos condiciones utilizadas ($\alpha_s = ,05$; $\alpha_s = ,10$) se corresponden con las utilizadas por Vriens, Wedel y Wilms (1996), Wedel y Steenkamp (1989) y especialmente por Hagerty (1985). Según este autor, la homogeneidad afecta al *sesgo* que se produce al agrupar los sujetos en conglomerados. Este sesgo está en función del grado de correlación entre las preferencias verdaderas de los consumidores. Cuanto más correlacionados están dos sujetos, más pequeño es este error al unirse en un mismo conglomerado. De modo que cuando los segmentos son homogéneos, el sesgo a la hora de estimar la verdadera pertenencia de los sujetos a cada grupo es pequeño. Vriens, Wedel y Wilms han encontrado que los métodos que permiten una adscripción a varios segmentos (como el basado en clases latentes) presentan un mejor rendimiento cuando los segmentos son difusos.

Similaridad de los segmentos. Para simular una situación con segmentos poco similares entre sí, las utilidades de cada segmento α_{xs} se han multiplicado por dos en la condición disímil, lo que ha provocado una mayor separación de los segmentos. Vriens, Wedel y Wilms (1996) encuentran que un incremento en la distancia entre los segmentos mejora el rendimiento de los métodos de segmentación conjunta.

MEDIDAS DE COMPARACIÓN

Los índices empleados han sido los siguientes:

El porcentaje de varianza que explican los modelos conjuntos: R^2 .

Un error típico de estimación (ETE) calculado entre las utilidades verdaderas de cada sujeto y las estimadas para ese mismo sujeto por cada modelo:

$$ETE = \sqrt{\frac{\sum_{x=1}^X \sum_{n=1}^N (\alpha_{xn} - \hat{\alpha}_{xs})^2}{X \cdot N}} \quad (6)$$

donde u_{xn} es la utilidad verdadera del sujeto n ; y \hat{u}_{xs} es la utilidad estimada del segmento al que pertenece el sujeto n .

La proporción de sujetos correctamente clasificados por los diferentes métodos de segmentación en sus verdaderos segmentos (%CLASOK).

Un error típico de predicción (ETP) calculado entre las preferencias observadas en cada sujeto (Y_{ni}) y las predichas por cada modelo (\hat{Y}_{si}) en la muestra de validación (los ocho perfiles *holdout* considerados):

$$ETP = \sqrt{\frac{\sum_{n=1}^N \sum_{i=1}^I (Y_{ni} - \hat{Y}_{si})^2}{N \cdot I}} \quad (7)$$

La proporción de primeras elecciones correctas en la muestra de validación (%1ªOK), es decir, la proporción de sujetos cuyo estímulo *holdout* más preferido es predicho correctamente por cada modelo.

Esta cinco medidas abarcan los principales aspectos a tener en cuenta de las soluciones conglomerativas que ofrecen los tres métodos de segmentación comparados: su bondad de ajuste (1), qué tal realizan la estimación de los parámetros (medidas 2 y 3) y cuál es su capacidad predictiva (4 y 5). Medidas similares a éstas han sido empleadas previamente de forma sistemática (Andrews, Ansari y Currim, 2002; DeSarbo, Oliver y Rangaswamy, 1989; Green y Helsén, 1989; Hagerty, 1985; Vriens, Wedel y Wilms, 1996; Wedel y Kistemaker, 1989; Wedel y Steenkamp, 1989, 1991).

En consecuencia, de cada sujeto se han obtenido medidas en 15 variables (cinco índices x tres métodos). Los datos han sido analizados mediante contrastes no paramétricos, dado que las variables consideradas no cumplían las condiciones necesarias para poder conocer la distribución muestral del estadístico F (en concreto, incumplían los supuestos de independencia, normalidad y, en su caso, aditividad). El factor *método de segmentación* ha sido contrastado mediante pruebas para muestras relacionadas (las soluciones conglomerativas se obtuvieron a partir de los mismos 600 sujetos). El resto de los factores considerados, mediante pruebas para muestras independientes (cada factor divide a la muestra en 300 sujetos diferentes).

RESULTADOS

Los resultados de los contrastes para los efectos principales aparecen recogidos en la Tabla 4.

Los tres métodos de segmentación analizados sólo se diferencian significativamente en su capacidad predictiva (ETP). Si bien es cierto que SPHU obtiene mejores resultados en todos los índices, sólo en su capacidad predictiva logra este método ser significativamente superior a SPHP y SCL, como se puede observar en la Tabla 5.

Tabla 4. Pruebas de Efectos Principales Sobre los Índices de Rendimiento (n = 600).

fuente (gl)	valores χ^2 ^a				
	R^2	ETE	%CLASOK	ETP	%1 ^o OK
método de segmentación (2)	,984 ^b (,611)	1,285 ^b (,526)	5,250 ^c (,072)	10,556 ^b (,005)	,400 ^c (,819)
nº de perfiles (1)	10,764 ^d (,001)	4,826 ^d (,028)	52,775 ^d (,000)	,003 ^d (,955)	,005 ^d (,974)
nº de segmentos (1)	3,896 ^d (,048)	13,325 ^d (,000)	37,702 ^d (,000)	3,188 ^d (,074)	21,806 ^d (,000)
varianza de error (1)	246,173 ^d (,000)	2,076 ^d (,150)	2,761 ^d (,097)	339,899 ^d (,000)	20,602 ^d (,000)
homogeneidad segmental (1)	,281 ^d (,596)	2,721 ^d (,099)	,350 ^d (,554)	,787 ^d (,375)	,001 ^d (,974)
separación segmental (1)	149,995 ^d (,000)	9,829 ^d (,002)	18,730 ^d (,000)	,001 ^d (,973)	15,459 ^d (,000)

a. Sus probabilidades asociadas aparecen, entre paréntesis, bajo cada valor.

b. ANOVA de Friedman.

c. Prueba Q de Cochran.

d. ANOVA de Kruskal-Wallis.

Menos relacionado con nuestros objetivos, aunque igualmente interesante, la Tabla 4 muestra también la significatividad estadística de varias características de los datos. En la Tabla 6 aparecen los valores medios de los diferentes índices bajo cada uno de los niveles considerados. Se observa, por ejemplo, que el ajuste de los modelos empeora a medida que disminuye el número de perfiles o la separación entre segmentos (segmentos difusos) y también cuando aumenta el número de segmentos o la varianza de error. Por su parte, la estimación de los parámetros se ve afectada por el número de perfiles, de segmentos y la separación entre los mismos. En la capacidad predictiva influye fundamentalmente la varianza de error.

Localizadas las diferencias entre los tres métodos en el ETP, un aspecto de interés puede ser intentar determinar dónde se localizan las mayores diferencias en la capacidad predictiva de los métodos en función de las

características de los datos consideradas en este análisis. Los resultados aparecen recogidos en la tabla de la página siguiente (Tabla 7).

Tabla 5. Rendimiento Medio de los Diferentes Métodos de Segmentación.

método	índices*				
	R ² ^a	ETE ^a	%CLASOK ^b	ETP ^a	%1 ^a OK ^b
1. SPHU	,8583	,7174	,9250	2,6579 ^{2,3}	,6117
2. SPHP	,8578	,7373	,9183	2,6827	,6117
3. SCL	,8569	,7434	,9167	2,6856	,6100

a. Diferencias entre cada par de métodos contrastadas mediante la prueba de los signos.

b. Diferencias contrastadas mediante la prueba de McNemar.

* Cada superíndice denota una superioridad estadísticamente significativa del método en que se encuentra sobre el que designa con su valor.

Tabla 6. Rendimiento Medio de los Diferentes Niveles Considerados.

factores	índices				
	R ²	ETE	%CLASOK	ETP	%1 ^a OK
perfiles					
9	,8364	,8083	,8422	2,7599	,6100
27	,8790*	,6486*	,9978*	2,5882	,6122
segmentos					
4	,8622*	,6666*	,9878*	2,5496	,7033*
8	,8531	,7935	,8522	2,7956	,5189
varianza de error					
5%	,9322*	,6768	,9333	1,7665*	,7000*
35%	,7831	,7848	,9067	3,3459	,5222
homogeneidad segmental					
$\chi^2 = ,05$,8613	,7365	,9122	2,6657	,6100
$\chi^2 = ,10$,8541	,7290	,9278	2,6851	,6122
separación segmental					
similares	,7927	,7656	,8756	2,6469	,5344
disímiles	,9226*	,6985*	,9644*	2,7036	,6878*
medias globales					
	,8577	,7328	,9200	2,6754	,6111

* Indica que, para cada índice, la diferencia entre los dos niveles del atributo correspondiente es significativa ($p < ,05$), contrastado mediante la prueba U de Mann-Whitney.

Tabla 7. Error de Predicción Medio (ETP) de Cada Método de Segmentación Según los Diferentes Niveles Considerados.

factores	métodos			
	1. SPHU	2. SPHP	3. SCL	Media
perfiles				
9	2,7304 ³	2,7717	2,7772	2,7599
27	2,5833	2,5907	2,5907	2,5882
segmentos				
4	2,5405	2,5514	2,5569	2,7956
8	2,7703 ³	2,8079 ³	2,8084	2,5496
varianza de error				
5%	1,7607 ²	1,7687	1,7702	1,7665
35%	3,3210 ³	3,3564	3,3602	3,3459
homogeneidad segmental				
$\eta^2 = ,05$	2,6349 ^{2,3}	2,6866	2,6753	2,6657
$\eta^2 = ,10$	2,6807	2,6788	2,6958	2,6851
separación segmental				
similares	2,6175 ^{2,3}	2,6598	2,6632	2,6469
disímiles	2,6977	2,7054	2,7078	2,7036

* Cada superíndice denota una superioridad estadísticamente significativa (contrastada mediante la prueba de los signos) del método en que se encuentra sobre el que designa con su valor.

Los tres métodos son muy similares entre sí cuando la muestra es pequeña, cuando los segmentos son pocos, muy heterogéneos y bien separados entre sí, y en situaciones experimentales con muchos grados de libertad. Pero a medida que aumenta el número de sujetos de la muestra, el número de segmentos, la varianza de error y disminuye los grados de libertad, la heterogeneidad y la separación de los segmentos (condiciones todas ellas más próximas a un estudio aplicado con datos reales), la solución conglomerativa que ofrece el procedimiento SPHU logra destacarse significativamente respecto al resto.

DISCUSIÓN Y CONCLUSIONES

El principal resultado de interés obtenido tras efectuar los análisis con datos simulados es la gran similitud observada entre las soluciones conglomerativas ofrecidas tanto por SPHU como por SPHP y SCL. Los tres métodos resultan ser enormemente coincidentes ... ¡hasta en los errores que cometen! (Figura 5):

	id	nºseg	sphp	sphu	scl
1	1	1	1	1	1
2	2	1	1	1	1
3	3	1	4	1	1
4	4	2	2	2	2
5	5	2	2	2	2
6	6	2	2	2	2
7	7	3	3	3	3
8	8	3	3	3	3
9	9	3	3	3	3
10	10	4	4	4	4
11	11	4	4	4	4
12	12	4	4	4	4
13	13	5	5	5	5
14	14	5	5	5	5
15	15	5	2	2	2
16	16	5	5	5	5
17	17	6	3	3	2
18	18	6	6	6	6
19	19	6	3	3	3
20	20	7	5	5	5
21	21	7	5	5	5
22	22	7	5	5	5
23	23	8	7	7	7
24	24	8	8	8	8
25	25	8	7	7	7

Figura 5. Verdaderas clasificaciones de los sujetos en los segmentos (nºseg) y estimaciones realizadas por los procedimientos SPHP, SPHU y SCL.

Sólo se detectan diferencias estadísticamente significativas en uno de los índices utilizados (*ETP*). En este caso, SPHU demuestra ser significativamente mejor que SPHP ($Z = -1,969$; $p = ,049$) y sobre todo mejor que SCL ($Z = -3,094$; $p = ,002$). Aunque sin llegar a la significatividad estadística, este mismo patrón se observa de hecho en casi todos los índices y características de los datos analizadas: una muy ligera pero consistente primacía de SPHU, seguida de SPHP y, en último lugar, SCL.

Estos resultados contrastan con los de Vriens, Wedel, y Wilms (1996), que encuentran una clara superioridad de los procedimientos de clases latentes sobre los métodos *post hoc*. Además de por haber utilizado diferentes condiciones experimentales (y no tratarse en consecuencia de un experimento réplica), otra posible explicación a esta discrepancia puede estar relacionada con el método conglomerativo utilizado. En nuestro estudio, SPHP y SPHU están basados en una metodología conglomerativa combinada (método jerárquico + método de optimización) que algunos autores (Ketchen y Shook, 1996; Picón, Varela, y Real, 2003; Punj y Stewart, 1983) consideran que logra mejorar la clasificación de los sujetos en los segmentos respecto a la

utilización de alguno de estos métodos por separado. De este modo, la utilización adecuada del análisis de conglomerados puede convertirse en una pieza clave para incrementar la efectividad de los procedimientos *post hoc* respecto a los de clases latentes (como afirman los propios DeSarbo, Wedel, Vriens y Ramaswamy, 1992). Un segundo aspecto que puede haber influido en la disparidad de resultados está relacionado con la posibilidad que ofrece el SCL de usar clasificaciones solapadas. Ya hemos comentado que en nuestro caso los sujetos han sido adscritos a un único segmento sin apenas merma informativa. Desconocemos si éste hubiese sido también el caso del trabajo de Vriens, Wedel, y Wilms (1996), pero parece claro que el uso de clasificaciones más complejas como las solapadas puede suponer una ventaja para la SCL a medida que aumenta el porcentaje de incertidumbre en la asignación. En cualquier caso, estos mismos autores han encontrado que el tipo de partición conglomerativa (no solapada, solapada, borrosa) no es tan relevante a la hora de evaluar el rendimiento de un método como otros factores (tipo de algoritmo utilizado en la estimación, etc.).

Por último, destacar el hecho de que a pesar de que varios autores han teorizado acerca de la potencial inestabilidad de las estimaciones realizadas a nivel individual (Kamakura, 1988; Wedel y Kistemaker, 1989; Wittink y Cattin, 1989), en este trabajo las diferentes soluciones conglomerativas basadas en las funciones de utilidad individuales (SPHU) han tenido un rendimiento similar (y en ocasiones hasta superior) que las basadas en estimaciones a nivel de segmento (SPHP). Curiosamente, tampoco DeSarbo, Wedel, Vriens y Ramaswamy (1992) han encontrado evidencia empírica que apoye esta extendida creencia en la inestabilidad de las estimaciones individuales.

ABSTRACT

A Monte Carlo comparison of three metric conjoint segmentation methods. This simulation study compares three metric conjoint segmentation methods. Two of them are two-stage procedures based on a combined hierarchical + iterative clustering methodology, while the other is an integrated procedure based on a latent class approach. Following Vriens, Wedel & Wilms (1996) we should expect the latent class procedure perform best. However, only small and mostly insignificant differences between the procedures are reported. The only statistical differences are found on predictive accuracy and favour the two-stage method based on individual utilities over the other two.

REFERENCIAS

- Andrews, R. L., Ansari, A. y Currim, I. S. (2002). Hierarchical Bayes versus finite mixture conjoint analysis models: A comparison of fit, prediction and partworth recovery. *Journal of Marketing Research*, 39, 87-98.
- Binder K. y Stauffer, D. (1984). A simple introduction to Monte Carlo simulation and some specialized topics. En K. Binder (Ed.), *Applications of the Monte Carlo method in statistical physics* (pp. 1-36). Berlín, Alemania: Springer-Verlag.
- Dempster, A. P., Laird, N. M. y Rubin, D. B. (1977). Maximum likelihood from incomplete data via the E-M algorithm. *Journal of the Royal Statistical Society*, 39(serie B), 1-38.
- DeSarbo, W. S. y Cron, W. L. (1988). A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification*, 5, 249-282.
- DeSarbo, W. S., Oliver, R. L. y Rangaswamy, A. (1989). A simulated annealing methodology for clusterwise linear regression. *Psychometrika*, 54, 707-736.
- DeSarbo, W. S., Wedel, M., Vriens, M. y Ramaswamy, V. (1992). Latent class metric conjoint analysis. *Marketing Letters*, 3, 273-288.
- Gentle, J. E. (1998). *Random number generation and Monte Carlo methods*. Nueva York, EUA: Springer-Verlag.
- Green, P. E., Carmone, F. J. y Wind, Y. (1972). Subjective evaluation models and conjoint measurement. *Behavioral Science*, 17, 288-299.
- Green, P. E., Carroll, J. D. y Goldberg, S. M. (1981). A general approach to product design optimization via conjoint analysis. *Journal of Marketing*, 45(3), 17-37.
- Green, P. E. y Helsen, K. (1989). Cross-validation assessment of alternatives to individual-level conjoint analysis: A case study. *Journal of Marketing Research*, 26, 346-350.
- Green, P. E. y Srinivasan, V. (1978). Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research*, 5, 103-123.
- Green, P. E. y Srinivasan, V. (1990). Conjoint analysis in marketing: New developments with implications for research and practice. *Journal of Marketing*, 54(4), 3-19.
- Hagerty, M. R. (1985). Improving the predicting power of conjoint analysis: The use of factor analysis and cluster analysis. *Journal of Marketing Research*, 22, 168-184.
- Kamakura, W. A. (1988). A least squares procedure for benefit segmentation with conjoint experiments. *Journal of Marketing Research*, 25, 157-167.
- Ketchen, D. y Shook, C. (1996). The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal*, 17, 441-458.
- Landau, D. P. y Binder, K. (2000). *A guide to Monte Carlo simulations in statistical physics*. Cambridge, Reino Unido: Cambridge University Press.
- Lewis, P. A. W. y Orav, E. J. (1989). *Simulation methodology for statisticians, operation analysts, and engineers*. Belmont, California, EUA: Wadsworth.
- Mooney, C. Z. (1997). *Monte Carlo simulation*. Sage University Papers series on Quantitative Applications in the Social Sciences, 07-116. Thousand Oaks, California, EUA: Sage.
- Pardo, A. y San Martín, R. (1998). *Análisis de datos en psicología II* (2ª ed.). Madrid: Pirámide.
- Picón, E. y Varela, J. (2000). Segmentando mercados con análisis conjunto: Una aplicación al sector turístico. *Psicothema*, 12 (supl. n°2), 453-458.
- Picón, E., Varela, J. y Real, E. (2003). Clasificación y segmentación *post hoc* con análisis de conglomerados. En J. Varela y J. P. Levy (Eds.), *Análisis multivariable para las ciencias sociales* (pp. 417-450). Madrid: Pearson Educación.
- Punj, G. y Stewart, D. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research*, 20, 134-148.

- Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo method*. Chichester, Nueva York, EUA: Wiley.
- Srinivasan, V. (1975). Linear programming computational procedures for ordinal regression. *Journal of the Association for Computing Machinery*, 23, 475-487.
- Stigler, S. M. (1978). Mathematical statistics in the early states. *Annals of Statistics*, 6, 239-265.
- Umesh, U. N. y Mishra, S. (1990). A Monte Carlo investigation of conjoint analysis index-of-fit: Goodness of fit, significance and power: *Psychometrika*, 55, 33-44.
- Varela, J. (2000). Modelización de atributos múltiples. Escalamiento multidimensional y análisis conjunto. En J. Martínez Gastey y otros (Coords.), *La investigación en marketing* (Tomo II, pp. 609-676). Barcelona: AEDEMO.
- Varela, J., Rial, A. y García, A. (2003). Análisis conjunto. En J. Varela y J. P. Levy (Eds.), *Análisis multivariable para las ciencias sociales* (pp. 507-566). Madrid: Pearson Educación.
- Vriens, M., Wedel, M. y Wilms, T. (1996). Metric conjoint segmentation methods: A Monte Carlo comparison. *Journal of Marketing Research*, 33, 73-85.
- Wedel, M. y Kamakura, W. (1998). *Market segmentation: Conceptual and methodological foundations*. Boston, Massachusetts, EUA: Kluwer Academic.
- Wedel, M. y Kistemaker, C. (1989). Consumer benefit segmentation using clusterwise linear regression. *International Journal of Research in Marketing*, 6, 45-49.
- Wedel, M. y Steenkamp, J.-B. E. M. (1989). Fuzzy clusterwise regression approach to benefit segmentation. *International Journal of Research in Marketing*, 6, 241-258.
- Wedel, M. y Steenkamp, J.-B. E. M. (1991). A clusterwise regression method for simultaneous fuzzy market structuring and benefit segmentation. *Journal of Marketing Research*, 28, 385-396.
- Wittink, D. R. y Cattin, P. (1981). Alternative estimation methods for conjoint analysis: A Monte Carlo study. *Journal of Marketing Research*, 18, 101-106.
- Wittink, D. R. y Cattin, P. (1989). Commercial use of conjoint analysis: An update. *Journal of Marketing*, 53(3), 91-96.

(Manuscrito recibido: 1 Septiembre 2003; aceptado: 23 Enero 2004)