Metodología de Encuestas
Vol 3, Núm 1, 2001, 75-95

MODELOS DE ESTRUCTURA LATENTE: UNA POTENTE HERRAMIENTA AL SERVICIO DEL ANÁLISIS DE ENCUESTAS

Marcelino Sánchez Rivero
Universidad de Extremadura

RESUMEN

El análisis de la información contenida en una encuesta se relaciona en muchas ocasiones con el deseo del investigador de identificar y cuantificar conceptos latentes o pseudolatentes, mediante técnicas de observación indirecta en las que un conjunto de variables observadas (o indicadores) se instrumentaliza para la medición de esos conceptos latentes. En este sentido, algunas de estas técnicas han alcanzado en los últimos años una gran notoriedad en el análisis estadístico de datos de encuestas, aunque sus orígenes se sitúen a principios de los años cincuenta. Estamos haciendo referencia, concretamente, al *análisis de estructura latente*, cuya aplicabilidad al análisis de encuestas es, desde nuestro punto de vista, incuestionable, pero que, sin embargo, no ha alcanzado todavía suficiente popularidad entre los investigadores sociales. En consecuencia, el objetivo del presente artículo es apuntar los aspectos metodológicos más relevantes del modelo de análisis de clases latentes general, del análisis de estructura latente simultáneo, del análisis multivariante latente y de los modelos con errores de medición alternativos a la tradicional escala de Guttman.

Palabras clave: variable latente, independencia local, restricciones, homogeneidad, escala de Guttman.

Introducción

En el ámbito de las ciencias sociales proliferan gran cantidad de conceptos y fenómenos cuya medida resulta tremendamente complicada como consecuencia de la dificultad que entraña su observación. Los conceptos de esta naturaleza, que pueden agruparse bajo la denominación genérica de conceptos *latentes*, requieren para su estudio de técnicas de observación indirecta, en las que la recolección de información de un conjunto de indicadores de estos fenómenos latentes puede facilitar el análisis de los mismos.

Para el estudio de este tipo de fenómenos existe una serie de técnicas estadísticas que, aunque han demostrado su utilidad en el análisis de encuestas, son muy poco conocidas por los investigadores sociales. El objeto del presente artículo es esbozar los aspectos metodológicos más destacados de estas técnicas e, indirectamente, promover el uso de las mismas entre los estudiosos de las ciencias sociales.

El análisis de clases latentes es una técnica estadística que clasifica a los individuos de una población en un conjunto de segmentos o clases de naturaleza exhaustiva y excluyente. Atendiendo a la clasificación tradicional de las técnicas de segmentación, el análisis de clases latentes es una técnica óptima[1] basada en criterios subjetivos específicos (es decir, en criterios relacionados con aspectos internos de los individuos, como actitudes, percepciones, preferencias y, en general, cualquier otro aspecto de naturaleza subjetiva).

Una de las principales ventajas de esta técnica de segmentación frente a las técnicas tradicionales es su carácter confirmatorio. Al igual que otras técnicas estadísticas empleadas para segmentar, como el análisis factorial o el análisis cluster, el análisis de clases latentes es un método exploratorio de poblaciones cuando no existe una teoría explícita sobre la naturaleza de los diferentes grupos o subpoblaciones en que puede ser dividida la misma. Sin embargo, por encima de su naturaleza exploratoria, el análisis de clases latentes permite realizar todo tipo de investigaciones confirmatorias sobre la naturaleza del concepto latente (como la propia existencia del concepto, la adecuación de los indicadores empleados para su estudio, la óptima distribución de la población en los segmentos identificados, el tamaño de cada segmento, el comportamiento de los individuos ubicados en cada segmento, etc.).

Por otro lado, las características latentes de una población pueden estar presentes en otra —u otras— población —o poblaciones—. De aquí que, en ocasiones, el investigador social podría estar interesado en comparar las estructuras latentes de estas dos (o más) poblaciones, pudiendo centrar su interés en aspectos tales como la

[1] Se dice que una técnica de segmentación es óptima cuando el número de segmentos no está previamente establecido, sino que trata de determinar cuál debe ser el número óptimo de los mismos, a diferencia de las técnicas de segmentación *a priori*, en las que el número de segmentos está prefijado.

verificación de que el número de segmentos de todas las poblaciones es o no el mismo, la contrastación de comportamientos similares de un mismo segmento en todas las poblaciones consideradas o, con carácter general, la prueba de cualquier hipótesis de homogeneidad (tanto parcial como completa) relativa a dichas poblaciones. Pues bien, todos estos análisis comparativos son posibles gracias a una generalización del análisis de clases latentes, conocido como análisis de estructura latente simultáneo. La aplicación de estos modelos exige que la variable latente analizada sea la misma en las poblaciones que se comparan y que el número de segmentos definidos en las mismas sea también coincidente puesto que, en caso contrario, las poblaciones serán completamente heterogéneas, con lo que el análisis comparativo carecerá de fundamento. Además, los tamaños muestrales respectivos deben ser suficientemente grandes, ya que, si no es así, podría suceder que, por falta de potencia estadística, no llegara a rechazarse en ningún caso la hipótesis de ausencia de diferencias entre las poblaciones comparadas.

La identificación de diferentes segmentos en una población puede realizarse atendiendo a un solo fenómeno latente o, por el contrario, considerando simultáneamente dos o más fenómenos latentes. De esta forma, es posible establecer múltiples clasificaciones de una misma población atendiendo a diferentes criterios, pudiéndose modelizar asimismo la asociación estadística existente entre todos los fenómenos latentes estudiados. Desde un punto de vista metodológico, todos estos objetivos pueden alcanzarse utilizando una extensión simple del análisis de clases latentes en la que la relación entre un conjunto de indicadores está explicada por más de una variable latente.

Por último, el análisis de escala ha sido tradicionalmente una de las técnicas estadísticas más empleadas en el estudio estadístico de encuestas. A pesar de que su utilización ha sido anecdótica frente a otras escalas (como la escala de Likert o la de diferencial semántico), la escala de Guttman se ha erigido en una herramienta muy útil en la medición de actitudes con el objetivo de estandarizar pruebas. Sin embargo, esta escala ha sido objeto de múltiples críticas, entre las que destaca la que hace referencia a su naturaleza determinística. Como respuesta a esta serie de críticas, han surgido en la literatura diversas alternativas probabilísticas a la escala de Guttman que, en buena medida, vienen a salvar los inconvenientes de ésta. Estos modelos no son más que casos especiales del análisis de clases latentes en el que se han impuesto restricciones específicas a determinados parámetros del modelo.

Es indiscutible, por tanto, la validez de los modelos de estructura latente para el análisis estadístico de encuestas. En los siguientes apartados se muestran los aspectos básicos de esta metodología, partiendo de un planteamiento general y continuando con determinados casos específicos que demuestran la multifuncionalidad de esta herramienta estadística en el ámbito de la investigación de encuestas.

El análisis de clases latentes general

Considérese el siguiente planteamiento inicial: un investigador social está interesado en profundizar en el conocimiento de las relaciones existentes entre, por ejemplo, cuatro variables de corte social de naturaleza categórica, que denominaremos *A*, *B*, *C* y *D*. En esta situación, es muy probable que este investigador desee encontrar otra variable, también categórica y que llamaremos *X*, que explique estas relaciones. Se dirá entonces que la variable *X* explica la relación entre las variables *A*, *B*, *C* y *D* cuando dicha relación desaparezca al fijar una categoría de la variable *X*, es decir, cuando las cuatro variables de partida se convierten en independientes de las demás al analizar la relación *condicionada* entre estas variables en la tabla de contingencia de dimensión cuatro $\{ABCD\}$ en *cada* categoría de la variable *X*.

En el escenario planteado, las variables *A*, *B*, *C* y *D* no están *directamente* relacionadas entre sí, pero cada una de ellas puede ser afectada por la variable *X*, y cuando la categoría o nivel de la variable *X* deja de ser fija, estos efectos entre *X* y las restantes cuatro variables pueden producir la *aparente* relación que existe entre estas últimas. Desde la óptica de la teoría de grafos, esta situación se representaría de la forma que consta en la figura 1.

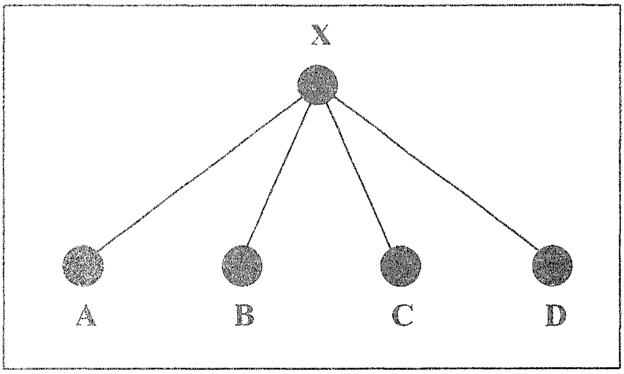

Figura 1: *La relación entre las variables A, B, C y D se debe a la relación entre éstas y X.*

Si la variable explicativa *X* es una variable que puede ser observada directamente, existen múltiples métodos para determinar si la misma explica o no adecuadamente las relaciones entre las variables *A*, *B*, *C* y *D*, entre los que destacan principalmente la modelización log-lineal, el modelo logit para variables categóricas, los modelos de asociación condicionada o los modelos de asociación parcial. El problema se plantea cuando la variable *X* no puede ser observada de una forma directa, es decir, cuando es una variable *inobservable* o *latente*. Éste es, precisamente, el origen del *Análisis de Clases Latentes* (ACL), que se concibe como una técnica estadística para la identificación de variables latentes y para la estimación de su influencia sobre un conjunto de variables observadas condicionalmente independientes de las mismas.

Al objeto de formular el modelo de clases latentes general, sea

$$\Pi_{ijklt}^{ABCDX}$$

la probabilidad conjunta de que un individuo se sitúe en las categorías i, j, k y l de las variables observadas A, B, C y D, respectivamente, y pertenezca a la clase t de la variable latente X. En virtud de la hipótesis que postula la figura 1, esta probabilidad conjunta puede descomponerse en el producto de una serie de factores:

$$\Pi_{ijklt}^{ABCDX} = \Pi_{it}^{\bar{A}X} \Pi_{jt}^{\bar{B}X} \Pi_{kt}^{\bar{C}X} \Pi_{lt}^{\bar{D}X} \Pi_{t}^{X} \quad (1)$$

donde

$$\Pi_{it}^{\bar{A}X}$$

denota la probabilidad condicionada de que un individuo, perteneciente a la clase t de la variable X, se sitúe en la categoría i de la variable observada A (las restantes probabilidades condicionadas,

$$\Pi_{jt}^{\bar{B}X}, \; \Pi_{kt}^{\bar{C}X} \; \text{y} \; \Pi_{lt}^{\bar{D}X}$$

se definen de forma similar y hacen referencia, respectivamente, a la categoría j de la variable B, a la categoría k de la variable C y a la categoría l de la variable D). Por su parte, el término

$$\Pi_{t}^{X}$$

representa la probabilidad de que un individuo pertenezca a la clase t de la variable latente X.

La probabilidad asociada a la casilla (i,j,k,l) de la tabla de contingencia obtenida a partir de la clasificación cruzada de las variables A, B, C y D, y que representaremos por

$$\Pi_{ijkl}^{ABCD},$$

se calcularía, a partir de la expresión (1), de la siguiente forma:

$$\Pi_{ijkl}^{ABCD} = \sum_{t=1}^{T} \Pi_{ijklt}^{ABCDX} = \sum_{t=1}^{T} \Pi_{it}^{\bar{A}X} \Pi_{jt}^{\bar{B}X} \Pi_{kt}^{\bar{C}X} \Pi_{lt}^{\bar{D}X} \Pi_{t}^{X} \quad (2)$$

La fórmula anterior es la del modelo de ACL *general* para el caso de cuatro variables observadas, que establece que *los individuos de una población pueden ser clasificados en T clases latentes (o segmentos) mutuamente exclusivas y exhaustivas*. Es en esta capacidad clasificatoria del análisis de clases latentes donde radica, precisamente, su carácter de *técnica de segmentación*. La asignación de los individuos a una clase latente o a otra se lleva a cabo calculando la siguiente probabilidad condicionada:

$$\Pi_{ijklt}^{ABCD\overline{X}} = \frac{\Pi_{ijklt}^{ABCDX}}{\Pi_{ijkl}^{ABCD}} \qquad (3)$$

En la expresión anterior,

$$\Pi_{ijklt}^{ABCD\overline{X}}$$

denota la probabilidad condicionada de que un individuo pertenezca a la clase latente t de la variable X, dado que se ha situado en el nivel (i,j,k,l) de la variable conjunta (A,B,C,D). Una vez calculadas estas T probabilidades condicionadas (correspondientes a las T clases de la variable X), los individuos con modalidad de respuesta (i,j,k,l) serán asignados a aquella clase latente para la cual se obtenga la probabilidad

$$\Pi_{ijklt}^{ABCD\overline{X}}$$

más alta. En consecuencia:

$$\text{Criterio de asignación de clases: } \max_{t} \Pi_{ijklt}^{ABCD\overline{X}}$$

Dado el carácter probabilístico de la asignación de individuos a las clases latentes de la variable X, siempre puede existir un cierto error en dicha asignación (este error sólo sería nulo cuando la probabilidad modal fuese igual a 1.00). De aquí que algunos autores hayan propuesto algunas medidas para cuantificar este error. Las más importantes son las propuestas por Clogg (1979, 1981) y se denominan *proporción correctamente clasificada* y *coeficiente* λ. Ambas medidas toman valores comprendidos entre 0 y 1, de forma que la asignación de los individuos a las clases de la variable latente será tanto más correcta cuanto más próximo a 1 se encuentre su valor. Estas medidas del error de asignación también pueden ser empleadas para cuantificar la asociación entre la variable latente y las variables observadas.

Por otro lado, en virtud de la naturaleza probabilística de los parámetros del modelo (2), se deberán verificar, además, las siguientes restricciones:

$$\sum_{i=1}^{I} \Pi_{it}^{\bar{A}X} = \sum_{j=1}^{J} \Pi_{jt}^{\bar{B}X} = \sum_{k=1}^{K} \Pi_{kt}^{\bar{C}X} = \sum_{l=1}^{L} \Pi_{lt}^{\bar{D}X} = 1 \tag{4}$$

$$\sum_{t=1}^{T} \Pi_{t}^{X} = 1 \tag{5}$$

La estimación de los parámetros del modelo de ACL general que, en nuestro caso, son las probabilidades condicionadas

$$\Pi_{it}^{\bar{A}X},\ \Pi_{jt}^{\bar{B}X},\ \Pi_{kt}^{\bar{C}X} \text{ y } \Pi_{lt}^{\bar{D}X}$$

y las probabilidades de clase latente

$$\Pi_{t}^{X}$$

se lleva a cabo mediante el método de la máxima verosimilitud. Este método demuestra que las estimaciones máximo-verosímiles de las probabilidades condicionadas y de las probabilidades de clase latente deben satisfacer un sistema de ecuaciones de verosimilitud, para cuya solución se utiliza un procedimiento iterativo, conocido con el nombre de *algoritmo EM*, propuesto originalmente por Goodman (1974a), y posteriormente desarrollado por Dempster, Laird y Rubin (1977). Básicamente, este algoritmo, que es una adaptación simple del método de Ajuste Proporcional Iterativo (o método IPF) empleado para estimar modelos log-lineales, consiste en partir de unos valores iniciales de las probabilidades condicionadas y de clase latente que satisfagan las ecuaciones y en obtener, en iteraciones sucesivas, nuevas estimaciones para estos parámetros. Este proceso iterativo de estimación finaliza cuando se llega a un número predeterminado de iteraciones o cuando se alcanza un *nivel de tolerancia* previamente fijado. Para una discusión más detallada sobre el algoritmo EM, puede consultarse Goodman (1974a) o Mooijaart y Van Der Heijden (1992).

Una vez obtenidas las estimaciones de los parámetros del modelo, es posible calcular las frecuencias esperadas estimadas de la tabla de contingencia bajo la hipótesis de que este modelo es cierto ($\hat{m}_{ijkl}$), de manera que para cuantificar la bondad de ajuste de dicho modelo a los datos bastará comparar éstas últimas con las correspondientes frecuencias observadas (n_{ijkl}), mediante el conocido test estadístico de la razón de verosimilitud:

$$G^2 = 2 \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{l=1}^{L} n_{ijkl} \log \frac{n_{ijkl}}{\hat{m}_{ijkl}} \tag{6}$$

que sigue una distribución ji-cuadrado asintótica con $IJKL$ - $[(I+J+K+L-3)T-1]$ grados de libertad.

Una de las grandes ventajas de esta técnica estadística es la posibilidad que ofrece de contrastar cualquier hipótesis relativa a las probabilidades condicionadas o a las probabilidades de clase latente, mediante la imposición de restricciones sobre las mismas. Surge de este modo el ACL *confirmatorio*, que permite al investigador social verificar todo tipo de conjeturas sobre la naturaleza de la variable latente identificada.

Las restricciones impuestas al modelo de ACL general pueden ser de igualdad o de valor. Las restricciones de igualdad tienen por objeto determinar si dos o más probabilidades (condicionadas o de clase latente) son idénticas. Mediante este primer tipo de restricciones es posible plantear, entre otras, hipótesis como las siguientes:

- La probabilidad de que un individuo se sitúe en la categoría i de la variable observada A es la misma para todas y cada una de las clases latentes de la variable X:

$$H_0 : \Pi_{i1}^{\bar{A}X} = \Pi_{i2}^{\bar{A}X} = \ldots = \Pi_{it}^{\bar{A}X} = \ldots = \Pi_{iT}^{\bar{A}X} \tag{7}$$

- La probabilidad de que un individuo pertenezca a la clase t de la variable X es la misma que la probabilidad de que pertenezca a la clase t' de dicha variable, es decir, las clases t y t' de la variable X poseen el mismo tamaño relativo (la proporción de individuos de la población pertenecientes a las clases t y t' es la misma):

$$H_0 : \Pi_t^X = \Pi_{t'}^X \quad para \;\; t \neq t' \tag{8}$$

- Las T clases de la variable X son equiprobables, esto es, las diferentes clases de la variable latente poseen todas el mismo tamaño relativo:

$$H_0 : \Pi_1^X = \Pi_2^X = \ldots = \Pi_t^X = \ldots = \Pi_T^X \tag{9}$$

Por su parte, las restricciones de valor tienen por objeto contrastar si una determinada probabilidad (tanto condicionada como de clase latente) coincide con un valor previamente determinado. Este segundo tipo de restricciones son especialmente útiles para valorar la importancia de una variable observada sobre las clases latentes o para ponderar el tamaño relativo de una clase latente concreta. Mediante las restricciones de valor, se pueden contrastar hipótesis como las que, a modo de ejemplo, se recogen a continuación:

- La probabilidad condicionada de que un individuo, que pertenece a la clase t de la variable X, se sitúe en la categoría j de la variable B es igual al 50 %:

$$H_0 : \Pi_{\mathrm{jt}}^{\overline{\mathrm{B}}\mathrm{X}} = 0{,}5 \tag{10}$$

- La probabilidad condicionada asociada a la categoría 3 de la variable C se sitúa en el límite de su espacio paramétrico:

$$H_0 : \Pi_{3\mathrm{t}}^{\overline{\mathrm{C}}\mathrm{X}} = 0 \text{ o bien } H_0 : \Pi_{3\mathrm{t}}^{\overline{\mathrm{C}}\mathrm{X}} = 1 \tag{11}$$

- Dentro de una clase latente determinada, las categorías de una variable observada son equiprobables:

$$H_0 : \Pi_{\mathrm{lt}}^{\overline{\mathrm{D}}\mathrm{X}} = \frac{1}{L} \quad para \quad \mathrm{l} = 1, 2, \ldots, L \tag{12}$$

Estas restricciones lineales determinan un único valor o un conjunto de valores de los parámetros dependiendo de que se trate de restricciones de valor o de restricciones de igualdad, respectivamente. La imposición de restricciones afecta al procedimiento de estimación por máxima verosimilitud de los parámetros del modelo, al fijarse uno o más de estos parámetros a priori, por lo que no será preciso estimarlos. Esta circunstancia determina que en el proceso de estimación se liberen tantos grados de libertad como parámetros estén restringidos, de forma que si el incremento en el valor del estadístico G^2 es pequeño en relación al aumento de los grados de libertad del modelo, se podrá admitir que las restricciones impuestas mejoran la bondad de ajuste del modelo sin restricciones a los datos observados. En caso contrario, las restricciones impuestas al modelo no restringido no podrían ser aceptadas estadísticamente.

El análisis de estructura latente simultáneo

Considérese ahora que se dispone de la clasificación cruzada de las variables observadas A, B, C y D para un total de S ($S \geq 2$) grupos o poblaciones diferentes. Esto significa la inclusión en el modelo de ACL de una variable grupal, que denotaremos por G, con categorías $s = 1, 2, \ldots, S$.

Sea

$$\Pi_{\mathrm{ijkls}}^{\overline{\mathrm{A}}\overline{\mathrm{B}}\overline{\mathrm{C}}\overline{\mathrm{D}}\mathrm{G}}$$

la probabilidad asociada a la combinación (i,j,k,l) de la variable conjunta (A,B,C,D) en el s-ésimo grupo considerado. Si se supone la existencia de un modelo de estructura latente de T clases para cada uno de los S grupos, se verificará que:

$$\Pi_{ijkls}^{\overline{A}\overline{B}\overline{C}\overline{D}G} = \sum_{t=1}^{T} \Pi_{ijklst}^{\overline{A}\overline{B}\overline{C}\overline{D}G\overline{X}} \tag{13}$$

donde:

$$\Pi_{ijklst}^{\overline{A}\overline{B}\overline{C}\overline{D}G\overline{X}} = \Pi_{ist}^{\overline{A}GX} \Pi_{jst}^{\overline{B}GX} \Pi_{kst}^{\overline{C}GX} \Pi_{lst}^{\overline{D}GX} \Pi_{st}^{G\overline{X}} \tag{14}$$

En las expresiones (13) y (14),

$$\Pi_{ijklst}^{\overline{A}\overline{B}\overline{C}\overline{D}G\overline{X}}$$

denota la probabilidad condicionada de que un individuo del s-ésimo grupo responda a las variables observadas A, B, C y D en las categorías i, j, k y l, respectivamente, y pertenezca a la clase t de la variable latente X. De forma similar, el término

$$\Pi_{st}^{G\overline{X}}$$

en (14) representa la probabilidad condicionada de que un individuo del grupo s pertenezca a la clase t de la variable X. Finalmente, el parámetro

$$\Pi_{ist}^{\overline{A}GX}$$

hace referencia a la probabilidad condicionada de observar la variable A en la categoría i cuando la variable X está en el nivel t para un individuo perteneciente al grupo o población s (las restantes probabilidades condicionadas,

$$\Pi_{jst}^{\overline{B}GX}, \ \Pi_{kst}^{\overline{C}GX} \text{ y } \Pi_{lst}^{\overline{D}GX}$$

de la expresión (14) se definen de forma similar a

$$\Pi_{ist}^{\overline{A}GX}).$$

Pues bien, combinando las expresiones (13) y (14) se obtendría la fórmula general del *modelo de estructura latente simultáneo de T clases*, que es la siguiente:

$$\Pi_{ijkls}^{\overline{A}\overline{B}\overline{C}\overline{D}G} = \sum_{t=1}^{T} \Pi_{ist}^{\overline{A}GX} \Pi_{jst}^{\overline{B}GX} \Pi_{kst}^{\overline{C}GX} \Pi_{lst}^{\overline{D}GX} \Pi_{st}^{G\overline{X}} \tag{15}$$

En la expresión anterior, las probabilidades de clase latente y las probabilidades condicionadas están sujetas a las siguientes restricciones:

$$\sum_{i=1}^{I} \Pi_{ist}^{\bar{A}GX} = \sum_{j=1}^{J} \Pi_{jst}^{\bar{B}GX} = \sum_{k=1}^{K} \Pi_{kst}^{\bar{C}GX} = \sum_{l=1}^{L} \Pi_{lst}^{\bar{D}GX} = 1 \qquad \forall s, t \tag{16}$$

$$\sum_{t=1}^{T} \Pi_{st}^{G\bar{X}} = 1 \qquad \forall s \tag{17}$$

Clogg y Goodman (1984) proponen una formulación alternativa para este modelo de estructura latente simultáneo mediante la definición de una nueva variable Y que recoja la clasificación cruzada de la variable grupal G y de la variable latente X (es decir, $Y = G\mathrm{x}X$) y la imposición de determinadas restricciones determinísticas sobre las probabilidades de clase latente y condicionadas del modelo. En definitiva, estos autores sostienen que un modelo de estructura latente de T clases para S grupos y $m = 4$ variables observadas puede ser realmente considerado como un modelo de ACL general con U clases (siendo $U = S\mathrm{x}T$) para una tabla de contingencia de dimensión cinco obtenida a partir de la clasificación cruzada de las cuatro variables observadas A, B, C y D y de la variable grupal G.

Sobre los parámetros del modelo (15) se pueden imponer diversos tipos de restricciones, siendo posible distinguir entre restricciones *intragrupales* y restricciones *intergrupales*. Aunque desde un punto de vista técnico no existen diferencias entre ambos tipos de restricciones, Clogg y Goodman (1984) hablan de *simples* restricciones para referirse a las restricciones intragrupales y de *restricciones de homogeneidad* para referirse a las restricciones intergrupales, ya que son éstas últimas las que poseen mayor interés metodológico.

Así, el modelo (15), en el que no se han impuesto restricciones de homogeneidad, se conoce con el nombre de *modelo de heterogeneidad completa*. Pero si se asume homogeneidad en todas las clases latentes del modelo, lo cual significa imponer las siguientes restricciones:

$$\Pi_{1t}^{G\bar{X}} = \Pi_{2t}^{G\bar{X}} = \ldots = \Pi_{st}^{G\bar{X}} = \ldots = \Pi_{St}^{G\bar{X}} = \Pi_{.t}^{G\bar{X}} \qquad \forall t \tag{18}$$

el modelo (15) se convertirá en un *modelo de homogeneidad parcial* (*en clases latentes*). También se pueden imponer restricciones de homogeneidad sobre las probabilidades condicionadas:

$$\Pi_{i1t}^{\bar{A}GX} = \Pi_{i2t}^{\bar{A}GX} = \ldots = \Pi_{ist}^{\bar{A}GX} = \ldots = \Pi_{iSt}^{\bar{A}GX} = \Pi_{i.t}^{\bar{A}GX} \qquad \forall i, t \tag{19}$$

$$\Pi_{j1t}^{\bar{B}GX} = \Pi_{j2t}^{\bar{B}GX} = \ldots = \Pi_{jst}^{\bar{B}GX} = \ldots = \Pi_{jSt}^{\bar{B}GX} = \Pi_{j.t}^{\bar{B}GX} \qquad \forall j, t \tag{20}$$

$$\Pi_{k1t}^{\bar{C}GX} = \Pi_{k2t}^{\bar{C}GX} = \ldots = \Pi_{kst}^{\bar{C}GX} = \ldots = \Pi_{kSt}^{\bar{C}GX} = \Pi_{k.t}^{\bar{C}GX} \qquad \forall k, t \tag{21}$$

$$\Pi_{l1t}^{\bar{D}GX} = \Pi_{l2t}^{\bar{D}GX} = \ldots = \Pi_{lst}^{\bar{D}GX} = \ldots = \Pi_{lSt}^{\bar{D}GX} = \Pi_{l.t}^{\bar{D}GX} \qquad \forall l, t \tag{22}$$

surgiendo de esta forma otros tantos modelos de homogeneidad parcial (*en probabilidades condicionadas*, en este caso).

Finalmente, un *modelo de homogeneidad completa* (tanto en clases latentes como en probabilidades condicionadas) podría obtenerse si se imponen las restricciones (18) a (22) sobre el modelo de heterogeneidad completa.

Modelos con múltiples variables latentes

En el planteamiento realizado hasta ahora se ha considerado que las relaciones de dependencia estadística entre las variables observadas A, B, C y D están explicadas por una única variable latente X. Sin embargo, podría suceder que no fuese una, sino varias, las variables latentes que expliquen la asociación existente entre las cuatro variables observadas anteriores.

A título ilustrativo, considérese que las variables observadas A y B son indicadores de una variable latente Y, y que las variables observadas C y D son indicadores de otra variable latente Z, de forma que las variables latentes Y y Z son estadísticamente dependientes. Esta situación puede representarse mediante el diagrama de la figura 2.

Sea

$$\Pi_r^Y$$

la probabilidad de que un individuo se sitúe en la clase r de la variable Y (r = 1, 2, ..., R); sea

$$\Pi_s^Z$$

la probabilidad de que un individuo pertenezca a la clase s de la variable Z (s = 1, 2, ..., S); y sea

$$\Pi_{rs}^{YZ}$$

la probabilidad de que un individuo se sitúe en el nivel (r,s) de la variable latente conjunta (Y,Z). Las probabilidades

$$\Pi_r^Y \text{ y } \Pi_s^Z$$

pueden obtenerse a partir de las probabilidades de clase latente conjuntas

$$\Pi_{rs}^{YZ}$$

de la siguiente forma:

$$\Pi_r^Y = \sum_{s=1}^{S} \Pi_{rs}^{YZ} \quad \forall r \tag{23}$$

$$\Pi_s^Z = \sum_{r=1}^{R} \Pi_{rs}^{YZ} \quad \forall s \tag{24}$$

Por otro lado, dado que un individuo pertenece a la clase (r,s) de la variable latente conjunta (Y,Z), se define

$$\Pi_{irs}^{\bar{A}YZ}$$

como la probabilidad condicionada de que responda en la categoría i de la variable A (las probabilidades condicionadas

$$\Pi_{jrs}^{\bar{B}YZ},\ \Pi_{krs}^{\bar{C}YZ} \text{ y } \Pi_{lrs}^{\bar{D}YZ},$$

referidas a las variables observadas B, C y D, respectivamente, se definen de forma similar).

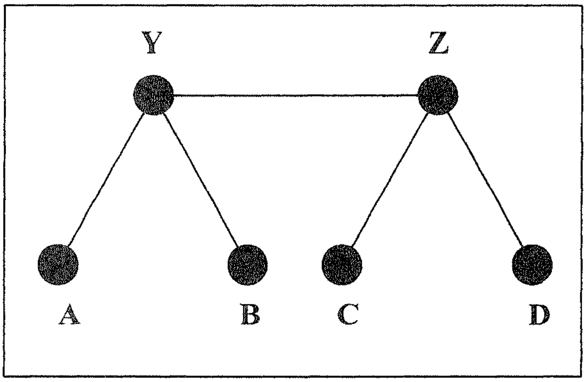

Figura 2: *Modelo con múltiples variables latentes.*

Pero puesto que, en virtud de la figura 2, las probabilidades condicionadas

$$\Pi_{irs}^{\bar{A}YZ} \text{ y } \Pi_{jrs}^{\bar{B}YZ}$$

dependen exclusivamente de la clase r de la variable Y (pero no de la clase s de la variable Z) y las probabilidades condicionadas

$$\Pi_{krs}^{\bar{C}YZ} \text{ y } \Pi_{lrs}^{\bar{D}YZ}$$

dependen únicamente de la clase s de la variable Z (pero no de la clase r de la variable Y), será necesario imponer las siguientes restricciones sobre dichas probabilidades condicionadas:

$$\Pi_{ir1}^{\bar{A}YZ} = \Pi_{ir2}^{\bar{A}YZ} = \ldots = \Pi_{irs}^{\bar{A}YZ} = \ldots = \Pi_{irS}^{\bar{A}YZ} = \Pi_{ir}^{\bar{A}Y} \quad \forall i, r \tag{25}$$

$$\Pi_{jr1}^{\bar{B}YZ} = \Pi_{jr2}^{\bar{B}YZ} = \ldots = \Pi_{jrs}^{\bar{B}YZ} = \ldots = \Pi_{jrS}^{\bar{B}YZ} = \Pi_{jr}^{\bar{B}Y} \quad \forall j, r \tag{26}$$

$$\Pi_{k1s}^{\bar{C}YZ} = \Pi_{k2s}^{\bar{C}YZ} = \ldots = \Pi_{krs}^{\bar{C}YZ} = \ldots = \Pi_{kRs}^{\bar{C}YZ} = \Pi_{ks}^{\bar{C}Z} \quad \forall k, s \tag{27}$$

$$\Pi_{l1s}^{\bar{D}YZ} = \Pi_{l2s}^{\bar{D}YZ} = \ldots = \Pi_{lrs}^{\bar{D}YZ} = \ldots = \Pi_{lRs}^{\bar{D}YZ} = \Pi_{ls}^{\bar{D}Z} \quad \forall l, s \tag{28}$$

Teniendo en cuenta las restricciones (25) a (28) y la figura 2, la expresión de un modelo de estructura latente con dos variables latentes (Y y Z) sería la siguiente:

$$\Pi_{ijkl}^{ABCD} = \sum_{r=1}^{R}\sum_{s=1}^{S} \Pi_{ijklrs}^{ABCDYZ} = \sum_{r=1}^{R}\sum_{s=1}^{S} \Pi_{ir}^{\bar{A}Y}\, \Pi_{jr}^{\bar{B}Y}\, \Pi_{ks}^{\bar{C}Z}\, \Pi_{ls}^{\bar{D}Z}\, \Pi_{rs}^{YZ} \tag{29}$$

La estimación de las probabilidades condicionadas y de las probabilidades de clase latente del anterior modelo se puede obtener también mediante el algoritmo EM, puesto que (29) puede reformularse como un modelo de estructura latente con una sola variable X que represente la distribución conjunta de las variables latentes Y y Z (es decir, $X = Y$xZ). Así, si la variable latente Y posee $R = 2$ clases y la variable latente Z posee $S = 3$ clases, el modelo de estructura latente con estas dos variables puede ser considerado como un modelo de ACL con 2x3 = 6 clases , en el que se verifican las siguientes restricciones sobre las probabilidades condicionadas[2]:

$$\Pi_{i1}^{\bar{A}X} = \Pi_{i2}^{\bar{A}X} = \Pi_{i3}^{\bar{A}X}\,;\; \Pi_{j1}^{\bar{B}X} = \Pi_{j2}^{\bar{B}X} = \Pi_{j3}^{\bar{B}X}\,;\; \forall i, j \tag{30}$$

$$\Pi_{i4}^{\bar{A}X} = \Pi_{i5}^{\bar{A}X} = \Pi_{i6}^{\bar{A}X}\,;\; \Pi_{j4}^{\bar{B}X} = \Pi_{j5}^{\bar{B}X} = \Pi_{j6}^{\bar{B}X}\,;\; \forall i, j \tag{31}$$

$$\Pi_{k1}^{\bar{C}X} = \Pi_{k4}^{\bar{C}X}\,;\; \Pi_{l1}^{\bar{D}X} = \Pi_{l4}^{\bar{D}X}\,;\; \forall k, l \tag{32}$$

$$\Pi_{k2}^{\bar{C}X} = \Pi_{k5}^{\bar{C}X}\,;\; \Pi_{l2}^{\bar{D}X} = \Pi_{l5}^{\bar{D}X}\,;\; \forall k, l \tag{33}$$

$$\Pi_{k3}^{\bar{C}X} = \Pi_{k6}^{\bar{C}X}\,;\; \Pi_{l3}^{\bar{D}X} = \Pi_{l6}^{\bar{D}X}\,;\; \forall k, l \tag{34}$$

[2] La correspondencia entre las clases (r,s) de la variable conjunta (Y,Z) y las clases t de la variable X sería, en este caso, la siguiente:

$(1,1) \to 1 \qquad (1,2) \to 2 \qquad (1,3) \to 3$

$(2,1) \to 4 \qquad (2,2) \to 5 \qquad (2,3) \to 6$

Finalmente, el planteamiento de un modelo de estructura latente con más de una variable latente exige el cumplimiento de una condición previa: la asociación estadística entre todas las variables latentes del mismo. Esto significa que si se formula un modelo log-lineal para analizar la dependencia estadística entre todas las variables latentes del modelo, al menos todos los efectos de interacción de segundo orden entre dichas variables deben ser estadísticamente distintos de cero. En caso contrario, el análisis puede realizarse de forma independiente mediante modelos de ACL con una única variable latente. Así, por ejemplo, si Y y Z fuesen estadísticamente independientes, el análisis estadístico de la variable Y podría realizarse formulando un modelo ACL en el que las variables observadas A y B serían indicadores de Y, mientras que el análisis estadístico de la variable Z se podría modelizar a través de un modelo ACL en el que los indicadores de esta variable latente serían las variables observadas C y D. Además, estos dos modelos ACL serían independientes, de forma que los resultados de uno no afectarían a la estimación y a la interpretación estadística del otro, y viceversa.

Para determinar si efectivamente las variables latentes del modelo son estadísticamente dependientes, es preciso estimar la magnitud de los efectos β^{YZ} y comprobar que éstos son estadísticamente significativos[3]. Para más detalles sobre el cálculo de estos efectos, puede consultarse Goodman (1974b).

Otras extensiones de los modelos de estructura latente

Es bien sabido que el análisis de escala se ha erigido en una valiosísima herramienta en la investigación social, ya que el uso de escalas unidimensionales, construidas a partir de un conjunto de ítems que representan un determinado concepto teórico, ha permitido realizar no sólo análisis exploratorios de datos, sino también otros de carácter confirmatorio.

Han sido muchas las escalas unidimensionales propuestas en la literatura para analizar el comportamiento de los individuos (escala de Thurstone, escala de Likert, etc.), pero al objeto de presentar otras posibilidades de los modelos de estructura latente, centraremos nuestra atención en la escala de Guttman. Este autor propuso en 1950 una escala unidimensional con k+1 posiciones, construida a partir de un con-

---

[3] En el caso de dos variables latentes con dos clases cada una, la estimación de β^{YZ}, que representa el efecto de interacción o de asociación entre dichas variables, viene dada por la siguiente expresión:

$$\hat{\beta}^{YZ} = \log\left[\sqrt{\frac{\hat{\Pi}_{11}^{YZ} \times \hat{\Pi}_{22}^{YZ}}{\hat{\Pi}_{12}^{YZ} \times \hat{\Pi}_{21}^{YZ}}}\right]$$

Si $\hat{\beta}^{YZ}$ es estadísticamente distinto de cero, se podrá admitir la dependencia estadística entre las variables latentes Y y Z.

junto de k ítems de naturaleza dicotómica (1 = respuesta afirmativa; 2 = respuesta negativa), de forma que la primera posición de la escala representa una respuesta afirmativa a todos los ítems, y la última posición de la escala una respuesta negativa a la totalidad de ítems, reflejando las posiciones intermedias, de forma gradual, respuestas negativas a los ítems empleados para construir la escala, los cuales están ordenados de menor a mayor dificultad de observación de una respuesta positiva.

Sin embargo, la escala de Guttman ha sido objeto de muchas críticas. Así, se le ha criticado su incapacidad para trabajar con grandes conjuntos de ítems, su asunción de que no existen datos perdidos e, incluso, su limitación a ítems de carácter dicotómico. Pero la crítica más importante es la que hace referencia a su naturaleza determinista, según la cual la probabilidad de respuesta a un ítem determinado de la escala sólo puede ser igual a 0 o igual a 1, excluyendo toda posibilidad de que la misma pueda variar entre estos dos valores extremos.

Es precisamente esta última crítica la que ha llevado a algunos autores a proponer alternativas probabilísticas a la escala de Guttman, que asumen la existencia de un error de medición comprendido entre 0 y 1. Pues bien, algunas de estas alternativas son extensiones del modelo de ACL, que se convierte de esta forma en una técnica para el análisis de escalabilidad de un conjunto de variables observadas.

Antes de presentar estas extensiones de los modelos de estructura latente, considérese, a título ilustrativo, un conjunto de 4 variables observadas o ítems dicotómicos, que denotaremos por A, B, C y D, respectivamente, y que éstos se emplean para construir una escala unidimensional de un fenómeno latente (variable X). Aunque, en este caso, existen un total de 16 modalidades de respuesta posibles una vez ordenados los 4 ítems de menor a mayor dificultad de observación de una respuesta positiva, el planteamiento de Guttman admite como válidas únicamente a 5 de estas modalidades de respuesta (1111, 1112, 1122, 1222, 2222), mientras que las 11 modalidades restantes son consideradas como modalidades de respuesta *incorrectas*.

A partir de estas condiciones iniciales, las principales alternativas probabilísticas a la escala de Guttman han sido las siguientes:

Modelo de Proctor:

Proctor (1970) considera que la escala unidimensional de Guttman posee porcentajes de error, que son homogéneos (iguales) para todos los ítems y para todas las posiciones de la escala. Los porcentajes de error vendrán dados por las correspondientes probabilidades condicionadas, dadas las diferentes posiciones de la escala. Para el caso que nos ocupa, las variables dicotómicas A, B, C y D definen una escala unidimensional con 5 posiciones, las cuales, a su vez, están representadas por 5 clases latentes en un modelo de ACL. La asunción de Proctor de homogeneidad de los porcentajes de error, tanto para los ítems como para las posiciones de escala, se representa mediante las siguientes restricciones de igualdad sobre las probabilidades condicionadas del modelo de ACL:

$$\Pi_{11}^{\bar{A}X} = \Pi_{11}^{\bar{B}X} = \Pi_{11}^{\bar{C}X} = \Pi_{11}^{\bar{D}X} = \Pi_{12}^{\bar{A}X} = \Pi_{12}^{\bar{B}X} = \Pi_{12}^{\bar{C}X} = \Pi_{22}^{\bar{D}X} =$$
$$\Pi_{13}^{\bar{A}X} = \Pi_{13}^{\bar{B}X} = \Pi_{23}^{\bar{C}X} = \Pi_{23}^{\bar{D}X} = \Pi_{14}^{\bar{A}X} = \Pi_{24}^{\bar{B}X} = \Pi_{24}^{\bar{C}X} = \Pi_{24}^{\bar{D}X} =$$
$$\Pi_{25}^{\bar{A}X} = \Pi_{25}^{\bar{B}X} = \Pi_{25}^{\bar{C}X} = \Pi_{25}^{\bar{D}X} \tag{35}$$

Como podrá apreciarse en la anterior expresión, este modelo requiere únicamente un grado de libertad para estimar las probabilidades condicionadas y 4 grados de libertad para estimar las probabilidades de clase latente.

Modelo de porcentaje de error propio de cada ítem

Ante la posibilidad de que no todos los ítems posean el mismo porcentaje de error, Clogg y Sawyer (1981) proponen un nuevo modelo en el que cada uno de los 4 ítems que forman la escala posee un porcentaje de error específico. El modelo de Clogg y Sawyer es también un caso particular del modelo de estructura latente general, en el que se han impuesto las siguientes restricciones sobre sus probabilidades condicionadas:

$$\begin{gathered}
\Pi_{11}^{\bar{A}X} = \Pi_{12}^{\bar{A}X} = \Pi_{13}^{\bar{A}X} = \Pi_{14}^{\bar{A}X} = \Pi_{25}^{\bar{A}X} \\
\Pi_{11}^{\bar{B}X} = \Pi_{12}^{\bar{B}X} = \Pi_{13}^{\bar{B}X} = \Pi_{24}^{\bar{B}X} = \Pi_{25}^{\bar{B}X} \\
\Pi_{11}^{\bar{C}X} = \Pi_{12}^{\bar{C}X} = \Pi_{23}^{\bar{C}X} = \Pi_{24}^{\bar{C}X} = \Pi_{25}^{\bar{C}X} \\
\Pi_{11}^{\bar{D}X} = \Pi_{22}^{\bar{D}X} = \Pi_{23}^{\bar{D}X} = \Pi_{24}^{\bar{D}X} = \Pi_{25}^{\bar{D}X}
\end{gathered} \tag{36}$$

En el caso que estamos considerando, esta alternativa probabilística a la escala de Guttman requiere 4 grados de libertad para estimar los porcentajes de error propios de cada ítem y otros 4 grados de libertad para estimar las probabilidades de clase latente del modelo.

Modelo de porcentaje de error propio de cada posición de la escala

Propuesto también por Clogg y Sawyer (1981), este modelo asume la existencia de porcentajes de error específicos de cada posición de la escala, hipótesis que se instrumenta mediante la imposición de las siguientes restricciones de igualdad sobre las probabilidades condicionadas del modelo de estructura latente general:

$$\begin{gathered}
\Pi_{11}^{\bar{A}X} = \Pi_{11}^{\bar{B}X} = \Pi_{11}^{\bar{C}X} = \Pi_{11}^{\bar{D}X} \\
\Pi_{12}^{\bar{A}X} = \Pi_{12}^{\bar{B}X} = \Pi_{12}^{\bar{C}X} = \Pi_{22}^{\bar{D}X} \\
\Pi_{13}^{\bar{A}X} = \Pi_{13}^{\bar{B}X} = \Pi_{23}^{\bar{C}X} = \Pi_{23}^{\bar{D}X} \\
\Pi_{14}^{\bar{A}X} = \Pi_{24}^{\bar{B}X} = \Pi_{24}^{\bar{C}X} = \Pi_{24}^{\bar{D}X}
\end{gathered}$$

$$\Pi_{25}^{\bar{A}X} = \Pi_{25}^{\bar{B}X} = \Pi_{25}^{\bar{C}X} = \Pi_{25}^{\bar{D}X} \quad (37)$$

Este modelo requiere 5 grados de libertad para estimar los porcentajes de error propios de cada clase latente o posición de escala, y 4 grados de libertad para estimar las probabilidades de clase latente.

Modelo de distancia latente de Lazarsfeld

Se trata de la alternativa probabilística a la escala de Guttman más elaborada, puesto que considera que el porcentaje de error para las respuestas afirmativas es diferente al porcentaje de error para las respuestas negativas. Esta distinción es aplicable a todos los ítems de la escala a excepción de los dos ítems extremos, para los que sólo existirá un porcentaje de error, válido tanto para las respuestas afirmativas como para las respuestas negativas (Lazarsfeld y Henry, 1968). Además de esta distinción entre los porcentajes de error para respuestas afirmativas y negativas, Lazarsfeld (1950) asume también que estos porcentajes son idénticos para todas las clases latentes, pero diferentes para cada ítem. Las restricciones que habría que imponer sobre las probabilidades condicionadas del modelo de ACL para obtener el modelo de distancia latente serían las siguientes:

$$\begin{gathered}
\Pi_{11}^{\bar{A}X} = \Pi_{12}^{\bar{A}X} = \Pi_{13}^{\bar{A}X} = \Pi_{14}^{\bar{A}X} = \Pi_{25}^{\bar{A}X} \\
\Pi_{11}^{\bar{B}X} = \Pi_{12}^{\bar{B}X} = \Pi_{13}^{\bar{B}X} \\
\Pi_{24}^{\bar{B}X} = \Pi_{25}^{\bar{B}X} \\
\Pi_{11}^{\bar{C}X} = \Pi_{12}^{\bar{C}X} \\
\Pi_{23}^{\bar{C}X} = \Pi_{24}^{\bar{C}X} = \Pi_{25}^{\bar{C}X} \\
\Pi_{11}^{\bar{D}X} = \Pi_{22}^{\bar{D}X} = \Pi_{23}^{\bar{D}X} = \Pi_{24}^{\bar{D}X} = \Pi_{25}^{\bar{D}X}
\end{gathered} \quad (38)$$

Como se deduce de la expresión (37), el modelo de distancia latente requiere la estimación de 6 probabilidades condicionadas y de 4 probabilidades de clase latente, por lo que el número de grados de libertad necesarios para estimar todos los parámetros del modelo asciende, en este caso, a 10.

Algunos ejemplos de aplicación de los modelos de estructura latente

Los modelos de estructura latente constituyen una línea de investigación aplicada que se viene desarrollando desde hace más de 10 años en el Departamento de Economía Aplicada y Organización de Empresas de la Universidad de Extremadura. Las principales aplicaciones de estas técnicas se han realizado, por tanto, en el ámbito de la economía y de la empresa. Sin ánimo de exhaustividad, se citan a continuación algunas de estas aplicaciones.

Fajardo y Sánchez (1992) utilizan el ACL general para estudiar el riesgo laboral de la población activa de la provincia de Badajoz. A partir de una muestra operativa de 3373 individuos, se identificaron los riesgos laborales más significativos de las diferentes clases de trabajadores según su propensión al riesgo en los sectores agrícola, industrial y servicios.

El atractivo turístico de Extremadura como variable latente ha sido también objeto de análisis por Fajardo y Sánchez (1994). Con la información aportada por una muestra aleatoria de 2384 turistas, los autores utilizan modelos de estructura latente para clasificar a la demanda turística extremeña en cuatro segmentos en función de los rasgos turísticos que más resaltan de Extremadura.

Pérez (2000a) utiliza los datos de la Encuesta de Presupuestos Familiares del año 1995 para cuantificar el error de medición en la movilidad de la renta, utilizando para ello un modelo de clases latentes y un modelo latente de Markov (utilizado habitualmente para analizar cambios en el tiempo).

Pérez y Fajardo (2000b) utilizan modelos de estructura latente para medir la lealtad de voto en los municipios extremeños en las elecciones a la Asamblea Regional de Extremadura y clasificarlos de acuerdo a los patrones de lealtad que presenten.

Pérez y otros (2000c) fundamenta la construcción de un índice de pobreza en el modelo de clases latentes. Considerando una muestra operativa de 6493 hogares del Panel de Hogares de la Unión Europea, clasifica a los hogares españoles en 3 grandes grupos (ricos, de bienestar medio y de bienestar bajo).

El análisis de estructura latente simultáneo ha sido utilizado por Sánchez (1998) para comparar las estructuras latentes de la población de turistas españoles menores de 45 años y la de los turistas con 45 o más años. Tras la identificación de 5 clases de turistas en ambas poblaciones (social, ecológico, pasivo, por entretenimiento puro y recreativo) y la imposición de determinadas restricciones de homogeneidad, se detectaron diferencias significativas entre ambas poblaciones en el tamaño relativo de los segmentos y en el comportamiento específico de los mismos. La base de datos empleada fue el estudio 2193 del Centro de Investigaciones Sociológicas sobre el comportamiento de los españoles ante las vacaciones. Este mismo estudio fue el utilizado por Sánchez (2000a) para presentar el ACL como técnica de segmentación de mercados.

Sánchez y otros (2000b) se basan en un modelo con dos variables latentes para clasificar a la población española según su espíritu democrático (demócratas y antidemócratas) y según su actitud ante el desempleo (sindicalistas y solidarios) y para confirmar la independencia estadística de estas dos variables latentes.

Finalmente, Sánchez (1999) analiza la escalabilidad del fraccionamiento vacacional de los españoles mediante modelos con errores de medición. Partiendo nuevamente del estudio 2193 del Centro de Investigaciones Sociológicas, se identificaron 5 posiciones de escala, de forma que en la primera posición de escala se sitúan los individuos con menor probabilidad de fraccionar sus vacaciones, mientras que en la última posición de escala se ubican los individuos con mayor predisposición a dividir sus vacaciones en dos o más períodos de tiempo.

Conclusiones

A lo largo del presente artículo se han abordado las notas metodológicas básicas de una técnica estadística muy útil en la investigación mediante encuestas: los modelos de estructura latente. Los comentarios que siguen pretenden ser un extracto de esta modelización estadística:

1º) Los modelos de variables latentes se basan en la idea de que la aparente relación existente entre un conjunto de variables observadas de naturaleza categórica está causada por una (o varias) variable latente. La relación estadística entre las variables observadas es de *independencia condicionada*, dada una categoría determinada de la variable latente, que también es de tipo categórico.

2º) Según el modelo de ACL general, toda población objeto de estudio puede ser dividida en un número (no determinado previamente) de segmentos excluyentes y exhaustivos.

3º) Los parámetros del modelo de ACL son las probabilidades de clase latente y las probabilidades condicionadas, para cuya estimación se emplea habitualmente un método iterativo de estimación por máxima verosimilitud denominado algoritmo EM. Las estimaciones finales obtenidas con este método deben satisfacer un sistema de ecuaciones de verosimilitud que se obtiene al maximizar la función de verosimilitud de la muestra.

4º) El investigador social puede llevar a cabo cualquier tipo de análisis confirmatorio relativo a la variable latente identificada, mediante la imposición de restricciones al modelo general, restricciones que pueden ser de igualdad o de valor específico.

5º) El análisis de estructura latente simultáneo permite comparar el mayor o menor grado de homogeneidad existente entre dos o más grupos o poblaciones independientes. Para ello, se pueden imponer restricciones de homogeneidad al modelo inicial (no restringido o heterogéneo) que generan nuevos modelos de homogeneidad parcial o de homogeneidad completa.

6º) Si se considera un conjunto de variables observadas categóricas, podría ocurrir que éstas fuesen indicadores no de una, sino de dos o más variables latentes. Ante esta situación, los modelos de estructura latente se pueden emplear para identificar las diferentes variables latentes, las cuales deben estar estadísticamente asociadas.

7º) El ACL es también una técnica estadística muy útil para el análisis de escalabilidad de un conjunto de ítems. Diversos autores (Proctor, Clogg, Sawyer, Lazarsfeld) han propuesto alternativas probabilísticas a la escala de Guttman. Estas modelizaciones alternativas son simples extensiones del modelo de ACL, en el que se han impuesto restricciones específicas sobre algunas de sus probabilidades condicionadas.

Referencias

Clogg, C.C. (1979): Some latent structure models for the analysis of Likert-type data. *Social Science Research*, vol. 8, pp. 287-301.

Clogg, C.C. (1981): New developments in latent structure analysis. D.M. Jackson y E.F. Borgotta. *Factor Analysis and Measurement*, pp. 215-246. Beverly Hills, CA: Sage.

Clogg, C.C. y Goodman, L.A. (1984): Latent structure analysis of a set of multidimensional contingency tables. *Journal of the American Statistical Association*, vol. 79, nº 388, pp. 762-771.

Clogg, C.C. y Sawyer, D.O. (1981): A comparison of alternative models for analyzing the scalability of response patterns. S. Leinhardt (ed.): Sociological Methodology, pp. 240-280. San Francisco: Josey-Bass.

Dempster, A.P.; Laird, N.M. y Rubin, D.B. (1977): Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, serie B, vol. 39, pp. 1-38.

Fajardo, M.A. y Sánchez, M. (1992): Estructuras latentes y salud laboral. *Actas de la VI Reunión Anual de Asepelt-España*, vol. I, pp. 17-28. Granada, 4 y 5 de Junio de 1992.

Fajardo, M.A. y Sánchez, M. (1994): El modelo de clases latentes y el atractivo turístico de Extremadura. *Actas de la VIII Reunión Anual de Asepelt-España*, vol. I, pp. 277-284. Palma de Mallorca, 2 y 3 de Junio de 1994.

Goodman, L.A. (1974a): Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, vol. 61, nº 2, pp. 215-231.

Goodman, L.A. (1974b): The analysis of systems of qualitative variables when some of the variables are unobservable. I. A modified latent structure approach. *American Journal of Sociology*, vol. 79, pp. 1179-1259.

Guttman, L.L. (1950): The basis for scalogram analysis. *Measurement and Prediction*. Edited by S.A. Stouffer, pp. 60-90. John Wiley & Sons, Inc., New York.

Lazarsfeld, P.F. y Henry, N.W. (1968): *Latent Structure Analysis*. Boston: Houghton Mifflin.

Lazarsfeld, P.F. (1950): The logical and mathematical foundation of latent structure analysis. *Measurement and Prediction*. Edited by S.A. Stouffer, pp. 362-412. John Wiley & Sons, Inc., New York.

Mooijaart, A. y Van der Heijden (1992): The EM algorithm for latent class analysis with equality constraints. *Psychometrika*, vol. 57, nº 2, pp. 261-269.

Pérez, J. (2000a): Measurement error in income mobility. Congreso "*Fighting poverty and inequality through tax-benefit reform: empirical approaches*" . Barcelona, 25 de Noviembre de 2000.

Pérez, J. y Fajardo, M.A. (2000b): Determinación de la lealtad de voto mediante un modelo de clases latentes. Documento de trabajo. Aceptado para su publicación en la revista *Estadística Española*.

Pérez, J. y otros (2000c): Cálculo de un indicador de pobreza mediante un modelo de clases latentes. *Actas de la XIV Reunión Anual de Asepelt-España*. Publicación electrónica en CD-Rom. Oviedo, 22 y 23 de Junio de 2000.

Proctor, C.H. (1970): A probabilistic formulation and statistical analysis of Guttman scaling. *Psychometrika*, vol. 35, pp. 73-78.

Sánchez, M. (1998): Modelización estadística de tablas de contingencia: aplicación al análisis de la demanda turística española. Tesis Doctoral. Universidad de Extremadura. Badajoz.

Sánchez, M. (1999): Modelo ACL y modelos con errores de medición: técnicas para el análisis de escalabilidad de variables categóricas. *Actas de la XIII Reunión Anual de Asepelt-España*. Publicación electrónica en CD-Rom. Burgos, 17 y 18 de Junio de 1999.

Sánchez, M. (2000a): El análisis de clases latentes como técnica de segmentación de mercados: el caso de la demanda turística nacional. *Actas del II Congreso Universitario de Turismo*, pp. 327-353. Benicassim, 21-23 de Abril de 1999.

Sánchez, M. y otros (2000b): Simultaneous analysis of latent variables: an extension of latent class models. *Actas de la VI Conferencia Cemapre "Mathematics Applied to Economics and Management"*. Pendiente de publicación. Lisboa, 5-7 de Junio de 2000.