

Un método alternativo para identificar los hogares en situación de pobreza

por

JESÚS PÉREZ MAYO(1)
MIGUEL A. FAJARDO CALDERA

Departamento de Economía Aplicada
Universidad de Extremadura

RESUMEN

En este trabajo se pretende construir un índice de pobreza a partir de la información proporcionada por los hogares españoles presentes en el ciclo de 1995 del Panel de Hogares de la Unión Europea sobre su situación económica. Concretamente, las variables usadas son aquellas relativas a la capacidad de satisfacer las necesidades del hogar así como la variable relativa a la capacidad de hacer frente a las necesidades habituales del hogar con los ingresos mensuales del mismo. Para realizar este análisis se utiliza un modelo de clases latentes, mediante el cual se clasifica a los individuos (en este caso, hogares) resumiendo la información proporcionada por el conjunto de variables elegido.

(1) Los datos utilizados en este trabajo pertenecen al Panel de Hogares de la Unión Europea y han sido proporcionados por el Instituto Nacional de Estadística. El autor es el único responsable de las estimaciones y resultados de este trabajo. En consecuencia, este trabajo no refleja las opiniones del Instituto Nacional de Estadística y todos los errores son responsabilidad del autor. Este trabajo ha sido mejorado gracias a las aportaciones y comentarios de un evaluador anónimo.

Palabras clave: pobreza, clases latentes, hogares.

Clasificación AMS: 62H30, 62P20, 91B82

1. INTRODUCCIÓN

Este trabajo pretende presentar una metodología alternativa para la determinación de los niveles de pobreza de la sociedad así como la identificación de los hogares en dicha situación concreta.

A partir de la definición de pobreza establecida por el Consejo Europeo en 1984, según la cual son pobres “aquellas personas, familias o grupos cuyos recursos (materiales, culturales y sociales) son tan limitados que les hacen quedar excluidos del modo de vida mínimo aceptable en el estado miembro en que habiten”, se contempla que es relativo y multidimensional. Es relativo, porque compara los recursos con los del resto de la sociedad y multidimensional al considerar distintos aspectos en el nivel de vida.

Sin embargo, tradicionalmente se mide dicho concepto multidimensional, el nivel de vida, a través de un instrumento unidimensional, la renta o el gasto. Es decir, en lugar de una medición directa se opta por una indirecta y, además, generalmente se confunde el medio, ingresos, con el fin, nivel de vida.

Mientras que esta metodología se basa en el supuesto de una fuerte relación entre el nivel de vida y la renta, Ringen (1988) cuestionó dicha hipótesis al constatar que no se cumplía. Por tanto, recomendó una medición directa para determinar la pobreza. En este trabajo, el modelo contemplado sigue este enfoque, es decir, no se utiliza la renta para identificar los hogares pobres, sino la información proporcionada por los hogares sobre la capacidad de satisfacer sus necesidades.

En el siguiente apartado revisaremos distintos métodos de medición de la pobreza, de forma general y terminaremos presentando la metodología que aplicaremos en este trabajo.

Después, en el apartado titulado “El modelo de clases latentes”, expondremos la teoría de dicho modelo así como el procedimiento de estimación para conocer la probabilidad de pertenecer a los distintos grupos.

Una vez presentado el modelo teórico, en la siguiente sección se mostrará la aplicación a los datos pertenecientes a la ola de 1995 del Panel de Hogares de la Unión Europea para España.

2. REVISIÓN DE LOS MÉTODOS DE MEDICIÓN DE LA POBREZA

Existen distintas maneras de medir la pobreza relacionadas sobre todo con el método utilizado para determinar el umbral o la línea de pobreza. Así, se compara el nivel de vida de cada individuo u hogar con el umbral para identificarlo como “pobre” o “no pobre”.

La definición de umbral utilizada ha ido variando a lo largo del tiempo y, además, está íntimamente relacionada con la idea de pobreza que se intenta medir.

Entre las distintas categorías, podemos encontrar:

- El presupuesto estándar: se basa en el trabajo inicial de Rowntree (1901). Consiste en la valoración del coste de un conjunto de bienes considerados necesarios según algunos juicios normativos por parte de los investigadores. Mientras que en el estudio de Rowntree antes citado sólo se hacía referencia a necesidades básicas, en los últimos trabajos que siguen esta línea, como Bradshaw (1993), se amplía la cesta de bienes.

- Relacionada con la anterior se encuentra la línea oficial de pobreza utilizada en los Estados Unidos. Orshansky (1965) calculó los costes de un presupuesto mínimo de alimentación para algunas familias de diferentes tamaños y construyó los umbrales de pobreza triplicando dichos costes.

- En la mayoría de las naciones industrializadas, el umbral se sitúa en un punto de la distribución de la renta equivalente. Por ejemplo, Eurostat ha tomado el 60% de la renta equivalente mediana como línea de pobreza.

- Otra línea de pobreza viene determinada administrativamente. En este caso, correspondería a la renta mínima aplicada en la legislación social o fiscal.

- El método subjetivo, en el cual la línea de pobreza se construye a partir de la opinión declarada por la población a partir de una pregunta parecida a la siguiente “¿Cuál es la renta mensual más baja para que un hogar como el suyo satisfaga todas sus necesidades?”. Esta aproximación al problema ha sido desarrollada, entre otros, por Van Praag *et al.* (1982) o Muffels *et al.* (1990).

- Finalmente, una posible manera de identificar a la población pobre reside en la utilización de indicadores sociales. El pionero de esta línea es Townsend (1979), quien construyó un índice de privación a partir de algunos indicadores no monetarios. El umbral de pobreza es el nivel de renta de los hogares con mayor privación. Sin embargo, otros estudios, siguiendo a Mack y Lansley (1985), pretenden medir directamente la pobreza según el nivel de vida mediante un conjunto de indicadores

no monetarios. En lo que respecta a España, Martínez y Ruiz-Huerta (1999, 2000) han sido los primeros en intentar cuantificar el porcentaje de hogares españoles en situación de pobreza mediante indicadores no monetarios.

En este trabajo, seguimos esta última línea de aproximación al problema de la medición de la pobreza. Por tanto, se intenta determinar directamente el nivel de vida de los hogares y, más tarde, identificar aquellos con el nivel inferior. No obstante, también en el análisis de la pobreza según este enfoque se presentan distintas estrategias para evaluar, ponderar y agregar los hogares, estrategias que se pueden revisar en Martínez y Ruiz-Huerta (1999) y Brandolini y D'Alessio (2000). Dentro de las metodologías presentadas en las citas anteriores, el método propuesto en este trabajo se inserta dentro del análisis estadístico multivariante. En concreto, debido al objetivo del trabajo, identificar distintos grupos en la población según su nivel de vida (variable categórica), y el tipo de datos utilizados (variables discretas), consideramos que el modelo de clases latentes es el más indicado para medir el nivel de vida. Por tanto, serán pobres aquellos hogares situados en el grupo de la población con menor nivel de vida.

3. EL MODELO DE CLASES LATENTES

Los modelos de clases latentes fueron introducidos por Lazarsfeld (1950) y Henry y Lazarsfeld (1968). Por otro lado, los problemas de estimación e identificación han sido tratados por Anderson (1954) y McHugh (1956). Goodman (1974) conectó estos modelos con la teoría moderna de las tablas de contingencia y finalmente, podemos citar a distintos autores que han desarrollado estas técnicas como Agresti (1982), Andersen (1993), Bartholomew (1987), Clogg (1993), McCutcheon (1987), entre otros.

Las relaciones de dependencia entre las variables categóricas de una tabla de contingencia en muchos casos están provocadas por la existencia de una asociación entre cada una de ellas y otra variable no observable directamente, llamada variable latente.

El modelo de clases latentes es una técnica estadística que permite estudiar la existencia de una (o varias) variable(s) latente(s) a partir de un conjunto de variables explicativas observadas y definir, a partir de sus clases, una clasificación o tipología de los individuos analizados. En el modelo de clases latentes, tanto las variables observadas como la variable latente se consideran variables categóricas con dos o más categorías, de manera que es necesario que la relación entre las variables manifiestas verifique dos hipótesis previas:

a) Relación simétrica: no existe una variable explicada por un conjunto de variables explicativas, sino que cada variable de la tabla de contingencia puede quedar explicada por y explicar el comportamiento de cualquier otra variable categórica de la tabla. En resumen, supone la ausencia de una variable categórica de interés que pretenda explicarse a través del resto de variables categóricas de la tabla.

b) Independencia local: supone que dentro de cada categoría de la variable latente, las variables observadas son estadísticamente independientes, es decir, que las variables de la tabla de contingencia son condicionadamente independientes dada una clase determinada de la clase latente.

El modelo de clases latentes puede parametrizarse de dos formas distintas: mediante probabilidades condicionadas o un modelo log-lineal.

Supongamos un conjunto de variables categóricas, A, B, C y D, con un número de categorías I, J, K y L respectivamente. Por tanto, tendremos una tabla de contingencia de dimensión $I \times J \times K \times L$. Por otro lado, sea X una variable latente con un total de T clases. Las ecuaciones básicas del modelo de clases latentes son:

$$\pi_{ijkl} = \sum_{t=1}^T \pi_{ijklt} , \quad [1]$$

donde

$$\pi_{ijklt} = \pi_t \pi_{ijk|t} = \pi_t \pi_{i|t} \pi_{j|t} \pi_{k|t} \pi_{l|t} . \quad [2]$$

Como vemos, se cumple la hipótesis de relación simétrica porque cada una de las variables consideradas depende sólo de la variable latente y, además, dentro de cada clase de ésta, las variables son estadísticamente independientes (hipótesis de independencia local).

Aquí, π_{ijklt} representa la probabilidad de estar en la celda (i, j, k, l, t) de la distribución conjunta ABCDX. Además, π_t es la probabilidad de pertenecer a la clase latente t y $\pi_{ijk|t}$ es la probabilidad de tener un patrón de respuesta concreta dado $X=t$. El resto de los parámetros π son probabilidades condicionadas.

Por tanto, los parámetros del modelo de clases latentes son las probabilidades condicionadas $\pi_{i|t}$, $\pi_{j|t}$, $\pi_{k|t}$, $\pi_{l|t}$ y las probabilidades de las clases latentes π_t , que estarán sometidas a las siguientes restricciones:

$$\sum_{i=1}^I \pi_{i|t} = \sum_{j=1}^J \pi_{j|t} = \sum_{k=1}^K \pi_{k|t} = \sum_{l=1}^L \pi_{l|t} = 1 , \quad [3]$$

$$\text{y } \sum_{t=1}^T \pi_t = 1.$$

El modelo de clases latentes expuesto antes es idéntico al modelo log-lineal {AX, BX, CX, DX}, cuya expresión es

$$\log m_{ijklt} = \mu_0 + \mu_i^A + \mu_j^B + \mu_k^C + \mu_l^D + \mu_t^X + \mu_{it}^{AX} + \mu_{jt}^{BX} + \mu_{kt}^{CX} + \mu_{lt}^{DX}, \quad [4]$$

en donde $m_{ijklt} = N\pi_{ijklt}$. La ecuación anterior, además de la media general y los términos de una variable, contiene sólo los términos de interacción entre la variable latente X y las variables manifiestas.

La obtención de las estimaciones máximo-verosímiles de los parámetros de los modelos de clases latentes es más complicada que para los modelos log-lineales donde se observen todas las variables. Son utilizados varios métodos de estimación, entre los cuales los más utilizados son el algoritmo de Newton-Raphson y el algoritmo EM (Dempster, Laird y Rubin, 1977). En nuestro trabajo, vamos a utilizar este último.

Dicho algoritmo es un procedimiento iterativo de estimación que consta de dos pasos. En el paso E(speranza) se calculan todos los valores esperados dados los valores observados y los “actuales” parámetros del modelo. En el paso M(aximización), se maximiza la función de verosimilitud de todos los datos a partir de los valores esperados calculados en el paso anterior. Esto implica el cálculo de estimaciones actualizadas de los parámetros del modelo como si no faltaran datos. Las iteraciones continúan hasta que se alcanza la convergencia.

Así, finalmente, podemos obtener las estimaciones máximo-verosímiles

$$\hat{\pi}_{i|t}, \hat{\pi}_{j|t}, \hat{\pi}_{k|t}, \hat{\pi}_{l|t} \text{ y } \hat{\pi}_t; \quad [5]$$

a partir de las que es posible calcular las probabilidades

$$\hat{\pi}_{ijklt} \text{ y } \hat{\pi}_{ijkl} = \sum_{t=1}^T \hat{\pi}_{ijklt}. \quad [6]$$

El siguiente paso en el análisis es asignar cada individuo a las diferentes clases de la variable latente X. Para ello se calcula la probabilidad condicionada de que un individuo que se sitúe en las categorías i, j, k y l de las variables manifiestas A, B, C y D, pertenezca a la clase t de la variable X de la siguiente manera:

$$\hat{\pi}_{t|ijkl} = \frac{\hat{\pi}_{ijklt}}{\sum_{t=1}^T \hat{\pi}_{ijklt}}. \quad [7]$$

Dada esta probabilidad, la regla de asignación es mediante la probabilidad modal, es decir, los individuos situados en la celda (i, j, k, l) de la tabla serán asignados a aquella clase latente cuya $\hat{\pi}_{t|ijkl}$ sea mayor. Como vemos se utiliza un proceso bayesiano para realizar dicha asignación.

4. APLICACIÓN DEL MODELO DE CLASES LATENTES A LA MEDICIÓN DE LA POBREZA

Los datos analizados en este trabajo provienen del Panel de Hogares de la Unión Europea para España y 1995. El Panel de Hogares de la Unión Europea (PHOGUE) es una encuesta armonizada en todos los países de la Unión Europea. Se inicia en 1994 con la participación de 12 países y una muestra de 70.000 hogares, de los que a España correspondían 8.000. En concreto, en el año 1995 la muestra para España consta de 6522 hogares. La unidad de análisis elegida es el hogar debido a que las variables consideradas, como veremos más tarde, se refieren a éste y no al individuo.

Las variables elegidas están contenidas en el epígrafe dedicado a la situación económica del hogar. Informan sobre la *capacidad* del hogar para satisfacer una serie de necesidades como una calefacción adecuada, disfrutar de vacaciones, renovar el vestuario o el mobiliario, comida nueva habitual, invitar a amigos, afrontar los pagos corrientes.

En particular, las variables seleccionadas son:

HF002: situación a fin de mes (6 categorías, más N.C.)

HF003: capacidad de permitirse una calefacción adecuada (2 categorías)

HF004: capacidad de permitirse vacaciones pagadas (2 categorías)

HF005: capacidad de renovar el mobiliario parcialmente (2 categorías)

HF006: capacidad de comprar prendas de vestir nuevas (2 categorías)

HF007: capacidad de hacer una comida (2 categorías)

HF008: capacidad de invitar a una copa o comida (2 categorías)

HF009: retrasos al pagar el alquiler (2 categorías)

HF010: retrasos al devolver préstamos relacionados con la vivienda (2 categorías)

HF011: retrasos en el pago de recibos relacionados con la vivienda (2 categorías)

HF012: retraso en el pago de compras aplazadas (2 categorías)

Antes de comenzar con el análisis, se observó qué porcentaje de no-respuesta presentaban estas variables. Sólo 29 hogares de la muestra inicial no han respondido a, al menos, una de las cuestiones. En consecuencia, se decidió eliminar dichos hogares de la muestra objeto del análisis y el tamaño muestral de nuestro estudio es de 6493 hogares.

El trabajo que aquí presentamos se ha desarrollado en tres etapas. En la primera, se consideraron todas las variables excepto la primera. Se aplica, por tanto, el modelo de clases latentes a un conjunto de variables dicotómicas que sólo informan si las necesidades se satisfacen. No incluyen ningún tipo de información sobre el grado de satisfacción.

Con este modelo, se va a contrastar si la población es homogénea para estas variables. Si no lo es, identificaremos distintos grupos caracterizados por diferentes grados de satisfacción de las necesidades.

Para realizar los contrastes y las estimaciones se utilizará el programa estadístico LEM, elaborado por Vermunt (1997).

Estadísticamente, la hipótesis de homogeneidad de la población se corresponde con el modelo de una clase latente según el cual se plantea la independencia de las variables. La aceptación de un modelo con más de una clase latente significa que la población puede dividirse en tantos grupos como clases tenga el modelo.

En la tabla anterior, la calidad de ajuste viene determinada por el estadístico L^2 junto a sus grados de libertad y su probabilidad asociada y por el estadístico BIC (Criterio de Información Bayesiano). Utilizaremos este último porque pequeñas diferencias pueden ser estadísticamente significativas dado un gran volumen de casos, como ocurre en nuestra muestra.

Tabla 1**MODELOS DE CLASES LATENTES APLICADOS AL GRUPO 1(2)**

<i>Modelo</i>	L^2	<i>gl</i>	<i>p</i>	<i>BIC</i>
Independencia	8215,7411	10	0,0000	-676,8591
Modelo con 2 clases	7523,1642	21	0,0000	-6987,2587
Modelo con 3 clases	774,0427	32	0,0000	-7925,4309
Modelo con 4 clases	516,4087	43	1,0000	-8086,5016
Modelo con 5 clases	416,0040	54	1,0000	-8090,3430
Modelo con 6 clases	357,9192	65	1,0000	-8051,8646

Fuente: Elaboración propia

El BIC es un estadístico bayesiano propuesto por Raftery (1986) para muestras grandes. Se obtiene a partir de la siguiente expresión:

$$BIC = L^2 - (gl) \log N \quad [8]$$

donde L^2 es la razón de log-verosimilitud, gl los grados de libertad asociados y N el tamaño muestral. La regla de selección del modelo es elegir el modelo con el menor valor del estadístico. Cuando el BIC es negativo, se prefiere la hipótesis nula frente al modelo saturado. Es decir, la hipótesis nula que se contrasta es la validez del modelo para explicar las relaciones entre las variables y, por el contrario la alternativa correspondería al modelo saturado. En este punto, es preciso comentar que éste último es el que considera todas las interacciones y en el ejemplo mostrado en la expresión [4], el correspondiente modelo saturado sería el modelo log-lineal {ABCDX} con las interacciones de tercer, cuarto y quinto orden no contempladas en dicha ecuación.

En este caso concreto, el modelo que debería ser elegido sería el que contempla 5 clases latentes, es decir, 5 grupos distintos en la población.

Una vez elegido el modelo, mostramos las estimaciones máximo-verosímiles de las probabilidades condicionadas y de clase latente. El análisis de éstas explica las razones del cambio del conjunto de variables considerado.

(2) El GRUPO 1 hace referencia al conjunto de datos obtenido considerando sólo las variables HF003, HF004, HF005, HF006, HF007, HF008, HF009, HF010, HF011, HF012.

Tabla 2**PROBABILIDADES CONDICIONADAS Y DE CLASE LATENTE**

Variable observada	Clases					
		1	2	3	4	5
HF003	1	0,0641	0,2345	0,8082	0,1228	0,1369
	2	0,9359	0,7655	0,1918	0,8772	0,8631
HF004	1	0,0284	0,2976	0,9929	0,1281	0,0659
	2	0,9716	0,7024	0,0071	0,8719	0,9341
HF005	1	0,0180	0,2436	0,8648	0,0786	0,0000
	2	0,9820	0,7564	0,1352	0,9214	1,0000
HF006	1	0,2458	0,9835	0,9994	0,7213	0,7097
	2	0,7542	0,0165	0,0006	0,2787	0,2903
HF007	1	0,5009	0,9942	1,0000	1,0000	0,9776
	2	0,4991	0,0058	0,0000	0,0000	0,0224
HF008	1	0,0427	0,9475	0,9977	0,7682	0,6044
	2	0,9573	0,0525	0,0023	0,2318	0,3956
HF009	1	0,1093	0,0000	0,0017	0,1989	0,0211
	2	0,8907	1,0000	0,9983	0,8011	0,9789
HF010	1	0,0552	0,0078	0,0044	0,3327	0,0115
	2	0,9448	0,9922	0,9956	0,6673	0,9885
HF011	1	0,2287	0,0122	0,0027	0,3701	0,0436
	2	0,7713	0,9878	0,9973	0,6299	0,9564
HF012	1	0,0812	0,0083	0,0036	0,3181	0,0037
	2	0,9188	0,9917	0,9964	0,6819	0,9963
Probabilidades de clase latente		0,0403	0,3780	0,3746	0,0337	0,1735

Fuente: Elaboración propia

Si seguimos un criterio práctico, variables discriminantes, podemos observar cómo no existe ninguna diferencia(3) importante entre las clases en las probabilidades condicionadas de las últimas cuatro variables. En consecuencia, en una segunda etapa se decide eliminar éstas y, por tanto, en la medición de la pobreza que se realiza en este estudio no se contemplan los retrasos en los diversos pagos. Además, esta similitud puede venir causada porque en ellas se utiliza el mismo código para decir que no hay retrasos y que no se puede aplicar dicha cuestión al hogar.

En consecuencia, veamos en la tabla 3 la bondad del ajuste para los modelos aplicados a este nuevo conjunto de variables.

Tabla 3

MODELOS DE CLASES LATENTES APLICADOS AL GRUPO 2(4)

<i>Modelo</i>	L^2	<i>gl</i>	<i>p</i>	<i>BIC</i>
Independencia	6922,2665	57	0,0000	6421,8931
Modelo con 2 clases	986,9381	50	0,0000	548,0141
Modelo con 3 clases	180,2749	43	0,0000	-197,1997
Modelo con 4 clases	86,5018	27	0,0000	-229,5235
Modelo con 5 clases	39,6594	29	0,0897	-214,9165

Fuente: Elaboración propia.

Aunque según el criterio del BIC debería ser elegido el modelo con 4 clases, el examen de las probabilidades condicionadas muestra que varios grupos presentan una estructura de probabilidades condicionadas muy parecidas. En consecuencia, dado que el propósito de este estudio es encontrar grupos diferenciados y no ligeramente diferentes, se selecciona el modelo con 3 clases latentes. Esta decisión se basa en el carácter eminentemente práctico de este trabajo. Se busca una identificación clara de los grupos y un estudio previo de las probabilidades condicionadas en el caso de cuatro clases mostró que venían motivadas por la aparición de dos clases con comportamiento similar a la segunda clase de la tabla 4 y cuyos tamaños suman aproximadamente el tamaño estimado de la clase antes citada.

(3) A pesar de que en las clases 1 y 4 los valores son menores, la misma categoría presenta la probabilidad condicionada modal en todas las clases.

(4) El GRUPO 1 hace referencia al conjunto de datos obtenido considerando sólo las variables HF003, HF004, HF005, HF006, HF007, HF008.

Además, se diferenciaban en ligeras variaciones de los porcentajes que, en ningún caso, hacían cambiar las características de ausencia o presencia de los mismos indicadores.

Así, los hogares españoles pueden dividirse en 3 grandes grupos. El primer grupo, que recoge al 40,13% de la población, se caracteriza por la capacidad de satisfacer todas las necesidades previstas en las variables. Es decir, este grupo está formado por hogares que poseen un nivel de bienestar elevado y que, por tanto, pueden calificarse como “ricos”.

Tabla 4

PROBABILIDADES CONDICIONADAS Y DE CLASE LATENTE

PROBABILIDADES CONDICIONALES DE CLASE LATENTE				
Variable observada		Clases		
		1	2	3
HF003	1	0,7877	0,1978	0,0924
	2	0,2123	0,8022	0,9076
HF004	1	0,9761	0,2222	0,0278
	2	0,0239	0,7778	0,9722
HF005	1	0,8509	0,1616	0,0055
	2	0,1491	0,8384	0,9945
HF006	1	0,9995	0,9394	0,4412
	2	0,0005	0,0606	0,5588
HF007	1	1,0000	0,9955	0,8009
	2	0,0000	0,0045	0,1991
HF008	1	0,9979	0,9070	0,2504
	2	0,0021	0,0930	0,7496
Probabilidades de clase latente		0,4013	0,4781	0,1206

Fuente: Elaboración propia

Por otro lado, el segundo grupo, formado por el 47,81% de la población, está compuesto por hogares que, en su mayoría, no pueden satisfacer necesidades como la calefacción adecuada en la vivienda, las vacaciones fuera de casa o la renovación parcial del mobiliario. Por el contrario, necesidades como la comida de

carne o pescado cada dos días o la compra de prendas de vestir nuevas sí se cumplen. Tal vez puede llamar la atención el hecho de que se tenga la capacidad de invitar a familiares o amigos a tomar una copa o una comida fuera de casa; sin embargo, hay que considerar la redacción de la pregunta ya que se quiere saber si los hogares son capaces de hacer frente a esta necesidad al menos una vez al mes. Este grupo estaría formado por hogares de “bienestar medio”.

Finalmente, el tercer grupo de los hogares, compuesto por el 12,06%, no puede hacer frente a ninguna necesidad, excepto la relacionada con la comida y parcialmente la del vestuario. La probabilidad de satisfacer esta última condicionada a la pertenencia a esta clase es del 50%. Esta clase podría ser calificada como “de bienestar bajo” o “de pobreza”. Además, este análisis nos muestra que a pesar de pertenecer a la clase de menor bienestar, es difícil que no se satisfaga una necesidad tan básica como la alimentación.

Un análisis complementario al anterior es el relativo a los posibles errores cometidos al asignar los individuos a una clase latente determinada, esto es, los llamados errores de mala clasificación.

En primer lugar, calculamos la proporción esperada de los errores de clasificación al asignar cada individuo a la clase cuya probabilidad sea mayor.

$$E = \sum_{ijkl} [1 - \max(\pi_{xijkl})] \hat{\pi}_{ijkl} \quad [9]$$

La proporción correspondiente a nuestro caso es del 0,1061, lo cual indica que se espera que los errores cometidos en la asignación sean muy pequeños.

Por otro lado, puede calcularse otro indicador de la calidad de la asignación a una clase concreta. Es una modificación del contraste λ de Goodman y Kruskal realizada por Clogg (1981) e indica la reducción del error o la mejora de la predicción al clasificar las observaciones en lugar de asignar todas a la clase modal. Se define como

$$\lambda = \frac{[1 - \max(\hat{\pi}_x)] - E}{1 - \max(\hat{\pi}_x)} \quad [10]$$

En este caso, su valor es del 79,67%, lo cual indica de nuevo que la predicción al asignar un individuo a una clase determinada mejora apreciablemente.

Una vez hecho este estudio, se consideró por parte de los investigadores introducir la variable HF002 dentro del conjunto de variables. El objetivo de este último análisis es incluir la percepción que los hogares tienen de cómo llegan a fin de mes

con los ingresos netos mensuales. De esta forma, al combinar la información sobre la satisfacción de las distintas necesidades con la situación a la que llegan a fin de mes se mejora la descripción de las características de los grupos.

El modelo, elegido según el BIC, sería el modelo con 4 clases. Sin embargo, estamos de nuevo ante la cuestión surgida en el primer análisis. A pesar de presentar el menor valor del contraste, la información que proporciona a la hora de identificar los grupos según el nivel de vida es poco clara. Aparecen dos clases con una interpretación muy similar y probabilidades condicionadas muy parecidas.

En consecuencia, en aras del objetivo práctico del trabajo, preferimos tomar el modelo que contempla tres grupos latentes. Además, este modelo continúa siendo preferido al saturado.

Tabla 5

MODELOS DE CLASES LATENTES APLICADOS AL GRUPO 3(5)

<i>Modelo</i>	L^2	<i>gl</i>	<i>p</i>	<i>BIC</i>
Independencia	10075,9762	372	0,0000	6810,3817
Modelo con 2 clases	1846,3531	360	0,0000	-1313,8997
Modelo con 3 clases	557,2841	348	0,0000	-2497,6270
Modelo con 4 clases	336,4477	336	0,4829	-2613,1215
Modelo con 5 clases	264,8917	324	0,9929	-2579,3358

Fuente: Elaboración propia

En la tabla siguiente, podemos observar las probabilidades para el modelo de tres clases.

De nuevo, aparecen tres grupos con porcentajes de pertenencia parecidos al modelo anterior. El primer grupo, que engloba al 46,38%, se caracteriza por llegar con dificultad a fin de mes con sus ingresos netos mensuales, no poder satisfacer las necesidades relacionadas con la calefacción, las vacaciones y la renovación del mobiliario y si tienen capacidad para las demás. Por tanto, esta clase podría llamarse de "cercaños a la pobreza" porque pueden satisfacer las necesidades más básicas, pero se espera que tengan dificultad para lograrlo con sus ingresos netos mensuales.

(5) El GRUPO 3 hace referencia al conjunto de datos obtenido considerando sólo las variables HF001, HF003, HF004, HF005, HF006, HF007, HF008.

Tabla 6**PROBABILIDADES CONDICIONADAS Y DE CLASE LATENTE**

PROBABILIDADES CONDICIONALES DE CLASE LATENTE				
Variable observada		Clases		
		1	2	3
HF002	1	0,1208	0,5388	0,0089
	2	0,2422	0,2768	0,0459
	3	0,4492	0,1601	0,2416
	4	0,1642	0,0174	0,4385
	5	0,0224	0,0000	0,2388
	6	0,0011	0,0070	0,0263
HF003	1	0,2392	0,0860	0,8197
	2	0,7608	0,9140	0,1803
HF004	1	0,3098	0,0220	0,9803
	2	0,6902	0,9780	0,0197
HF005	1	0,2058	0,0076	0,8939
	2	0,7942	0,9924	0,1061
HF006	1	0,9655	0,5404	1,0000
	2	0,0345	0,4596	0,0000
HF007	1	0,9963	0,8605	1,0000
	2	0,0037	0,1395	0,0000
HF008	1	0,9355	0,4011	0,9979
	2	0,0645	0,5989	0,0021
Probabilidades de clase latente		0,4638	0,1753	0,3609

Fuente: Elaboración propia

Otra de las clases, compuesta por el 17,53%, es claramente la que contiene a los hogares en situación de “pobreza” ya que llegan con gran dificultades a fin de mes y sólo satisfacen las necesidades más básicas como la comida y el vestuario. De nuevo, llegamos a la misma conclusión que en el apartado anterior. La capaci-

dad de hacer una comida de carne y pescado cada dos días se espera que siempre se satisfaga.

Finalmente, el 36,09% de los hogares está recogido en una clase donde aparecen los hogares con más bienestar, ya que llegan con facilidad a fin de mes (como mínimo) y pueden satisfacer todas las necesidades.

Si, de nuevo, calculamos los estadísticos E y λ , se observa que el error cometido al asignar los hogares a una clase concreta es muy pequeño. Los valores son $E = 0,1054$ y $\lambda = 0,8035$

5. CONCLUSIONES

En resumen, a partir del estudio podemos llegar a esta serie de conclusiones:

1. Se ha planteado un método cualitativo, objetivo y de necesidades básicas para medir la pobreza al estudiar la homogeneidad de la población para una serie de variables.

2. El modelo de clases latentes puede ser utilizado para realizar dicho análisis si las variables disponibles en la encuesta son de naturaleza categórica.

3. Las variables relacionadas con el retraso en los pagos del alquiler, devolución de préstamos para la adquisición de la vivienda, recibos de la vivienda o compras aplazadas no discriminan entre los distintos grupos de hogares. Por tanto, es mejor realizar el estudio sin considerar estas variables.

4. Al analizar los resultados para el grupo 2 de variables, el modelo con 5 clases latentes debería ser elegido al presentar el menor BIC. No obstante, el modelo significativo más adecuado a los objetivos del estudio es el que contempla 3 clases latentes.

5. La principal conclusión es que se estima que los hogares, en su mayoría, tiene capacidad para satisfacer la necesidad de la alimentación.

6. Podemos dividir al conjunto de los hogares en tres grupos:

- Pobres: 12,06%
- Bienestar medio: 47,81%
- Bienestar alto: 40,13%.

Por este último, se entiende la capacidad de satisfacer todas las necesidades.

7. Si se incluye la información sobre la situación a la que llegan a fin de mes los hogares, de nuevo, el modelo más válido para nuestros propósitos es el que divide

la población en tres grupos con porcentajes de pertenencia parecidos. Se observa cómo los hogares antes calificados como pobres llegan con gran dificultad a fin de mes, los antes calificados como de bienestar medio llegan con dificultad y, finalmente, los de mayor bienestar llegan con gran facilidad.

8. Por tanto, proponemos la variable HF002 como aproximación observable al nivel de pobreza (latente).

REFERENCIAS

- ABUL NAGA, R.H. (1994): «Identifying the poor: a multiple indicator approach». Discussion Paper DARP 9. STICERD, London School of Economics.
- AGRESTI, A. (1982): «Analysis of ordinal categorical data». John Wiley and Sons, Nueva York.
- ANDERSEN, E.B. (1993): «The analysis of categorical data». Springer-Verlag, Berlín.
- ANDERSON, T.W. (1954): «On estimation of parameters in latent structure analysis». *Psychometrika* 19, 1-10.
- ATKINSON, A.B. ET AL (1997): «Measurement of trends in poverty and the income distribution». DAE Working Papers. Departamento de Economía Aplicada, Universidad de Cambridge.
- BARTHOLOMEW, D.J. (1987): «Latent variable models and factor analysis». Griffin, Londres.
- BRADSHAW, J. (ed.) (1993): «Budget Standards for the United Kingdom». Avebury, Aldershot.
- BRANDOLINI, A. Y D'ALESSIO, G. (2000): «Measuring well-being in the functionings space». 26ª Conferencia General de la IARIW (Asociación para el estudio de la Renta y la Riqueza). Cracovia, 27/08-02/09.
- CLOGG, C.C. (1993): «Latent class models: recent developments and prospects for the future». En ARMINGER, G., C.C. CLOGG Y M.E. SOBEL (1993): «Handbook of statistical modelling in the social sciences. Plenum, Nueva York.
- CREEDY, J. (1998): «The dynamics of inequality and poverty: comparing income distributions». Edward Elgar, Cheltenham
- DEATON A. (1997): «The analysis of household surveys: a microeconomic approach to development policy». The Johns Hopkins University Press, Baltimore.

- DEMPSTER, A.P., LAIRD, N.M. Y RUBIN, D.B. (1977): «Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion)». *Journal of the Royal Statistical Society, Serie B*, 39, 1-38.
- GOODMAN, L.A. (1974): «Exploratory latent structure analysis using both identifiable and unidentifiable models» *Biometrika*, 61, 215-231.
- HAGENAARS, J.A. (1990): «Categorical longitudinal data- loglinear analysis of panel, trend and cohort data». Sage, Newbury Park.
- HENRY, L.W. Y LAZARSFELD, P.F. (1968): «Latent structure analysis». Houghton Mifflin and Co., Boston.
- LAZARSFELD, P.F. (1950): «The logical and mathematical foundation of latent structure analysis», en STOUFFER ET AL, (1950): «Measurement and prediction». Princeton University Press, Princeton.
- MACK, J. Y LANSLEY, S. (1985): «Poor Britain». Allen & Unwin, Londres.
- MARTÍNEZ, R. Y RUIZ-HUERTA, J. (1999): «Algunas reflexiones sobre la medición de la pobreza. Una aplicación al caso español». En: MARAVALL, J.M. (ed.): «Dimensiones de la desigualdad. III Simposio sobre igualdad y distribución de la renta y la riqueza». Vol. 1, 367-428. Fundación Argentaria y Visor Editorial, Madrid.
- MARTÍNEZ, R. Y RUIZ-HUERTA, J. (2000): «Income, multiple deprivation and poverty: an empirical analysis using Spanish data». *26ª Conferencia General de la IAR-IW (Asociación para el estudio de la Renta y la Riqueza)*. Cracovia, 27/08-02/09.
- MCCUTCHEON, A.L. (1987): «Latent Class Analysis» Sage, Beverly Hills.
- McHUGH, R.B. (1956): «Efficient estimation and local identification in latent class analysis». *Psychometrika*, 21, 331-347.
- MUFFELS, R., FOUARGE, D. DEKKER, R. (1999): «Longitudinal poverty and income inequality. A comparative panel study for the Netherlands, Germany and the UK», WORC/TISSER Working paper. Universidad de Tilburg.
- MUFFELS, R., KAPTEYN, A. Y BERGHMAN, J. (1990): «Poverty in the Netherlands». VUGA's-Gravenhage.
- NUSSBAUM, M.C., SENA K. (1996): «La calidad de vida. Fondo de Cultura Económico», México.
- ORSHANSKY, M. (1965): «Counting the poor: another look at the poverty profile». *Social Security Bulletin*, Junio 3-29.

- RAFTERY A.E. (1986): «Choosing models for cross-classifications». *American Sociological Review*, 51-1, 145-146.
- RINGEN, S. (1988): «Direct and Indirect Measures of Poverty», *Journal of Social Policy*, 17(3), 147-164.
- ROWNTREE, S.B. (1901): «Poverty: A Study of Town Life». Longmans, Londres.
- SLOTTJE, D.J., RAJ, B. (1998): «Income inequality, poverty and economic welfare». Physica-Verlag, Heidelberg-Nueva York.
- TINBERGEN J. (1991), «On the measurement of welfare», *Journal of Econometrics*, North-Holland.
- TOWNSEND, P. (1979): «Poverty in the United Kingdom». Allen Lane, Hammondsworth.
- ULLAH, A., GILES, D.E.A. (1998): «Handbook of applied economic statistics». Marcel Dekker, Nueva York.
- VERMUNT, J.K. (1997): «Log-linear models for event histories». Sage, Londres.
- VAN PRAAG, B.M., HAGENAARS, A.J. Y VAN WEEREN, H. (1982): «Poverty in Europe». *Review of Income and Wealth*, 28: 345-359.
- VON EYE A., CLOGG C.C. (eds.) (1994) «Latent variables analysis: applications for developmental research». Sage, Londres.
- XIE, Y. (1992): «The log-multiplicative layer effect model for comparing mobility tables», *American Sociological Review*, 57-3, 380-395.

AN ALTERNATIVE MODEL FOR IDENTIFYING POOR HOUSEHOLDS

SUMMARY

A poverty index is built from the information on financial situation given by Spanish households in the 1995 wave of the European Community Household Panel. Exactly, the variables related to the capability of satisfying household needs and the one about economic strain are used. This analysis is performed by using a latent class model and so, individuals (households, in our case) are classified according the information given by the chosen variable set.

Keywords: poverty, latent class, households.

AMS Classification: 62H30, 62P20, 91B82