



Revista de Matemática: Teoría y Aplicaciones  
ISSN: 1409-2433  
[mta.cimpa@ucr.ac.cr](mailto:mta.cimpa@ucr.ac.cr)  
Universidad de Costa Rica  
Costa Rica

Araya A., Carlomagno; Sepulveda C., Rosa  
Análisis de clases latentes en tablas poco ocupadas: consumo de alcohol, tabaco y otras drogas en adolescentes  
Revista de Matemática: Teoría y Aplicaciones, vol. 17, núm. 1, 2010, pp. 25-40  
Universidad de Costa Rica  
San José, Costa Rica

Disponible en: <http://www.redalyc.org/articulo.oa?id=45326934003>

- ▶ Cómo citar el artículo
- ▶ Número completo
- ▶ Más información del artículo
- ▶ Página de la revista en redalyc.org

ANÁLISIS DE CLASES LATENTES EN TABLAS  
POCO OCUPADAS: CONSUMO DE ALCOHOL,  
TABACO Y OTRAS DROGAS EN ADOLESCENTES

LATENT CLASS ANALYSIS IN SPARSE DATA:  
ALCOHOL, TOBACCO AND OTHER DRUGS IN  
TEENAGERS

CARLOMAGNO ARAYA A.\*      ROSA SEPÚLVEDA C.\*\*

*Received: 28 Nov 2008; Revised: 28 Oct 2009;  
Accepted: 29 Oct 2009*

---

**Palabras clave:** clases latentes, adolescencia, drogas.

**Keywords:** latent class, adolescence, drugs.

**Mathematics Subject Classification:** 62H30, 62H17.

---

\*Sede de Occidente, Universidad de Costa Rica, San Ramón, Costa Rica. E-Mail:  
[cmaraya@cariari.ucr.ac.cr](mailto:cmaraya@cariari.ucr.ac.cr), [cmastat@gmail.com](mailto:cmastat@gmail.com)

\*\*Departamento de Estadística, Universidad de Salamanca, España. E-Mail:  
[a108813@usal.es](mailto:a108813@usal.es)

### Resumen

La aportación de este estudio es la aproximación multidimensional sobre patrones de consumo de drogas en jóvenes. Los datos provienen del estudio “*Factores que inciden en el consumo de drogas, población juvenil. Región Central de Occidente*” de Costa Rica, 2006. Con base en los resultados obtenidos con un modelo de clases latentes, se establecen 8 subgrupos de individuos según el consumo de las diferentes drogas.

### Abstract

The contribution of this study is the multidimensional approach on patterns of drug use among young people. The data come from the study ”Factors that influence the consumption of drugs, juvenile population. Central region of the West” from Costa Rica, 2006. On the basis of the results obtained with a latent class model, 8 subgroups of individuals settle down according to the consumption of different drugs.

## 1 Introducción

En la sociedad costarricense desde muy corta edad se da un alto consumo de las llamadas drogas lícitas, como es el caso del alcohol y el tabaco. No es de sorprender que la adolescencia sea la época más frecuente de experimentación con las drogas. Estas drogas se consideran drogas de “entrada” pues en muchos casos su consumo puede conducir en algún momento al uso de drogas ilícitas como la marihuana, la cocaína, etc. La edad promedio de inicio de consumo es de 13 años para el alcohol y de 13,3 años para el tabaco [16].

En el contexto nacional se presentan ciertos factores o circunstancias que universalmente se han establecido como favorecedores del consumo de drogas. Entre ellos, se pueden destacar la comunicación deficitaria entre padres e hijos, diversos tipos de carencias afectivas, y un clima social que no favorece el crecimiento personal, con falta de independencia y de capacidad para afrontar problemas personales, generándose por consiguiente una clara y directa relación entre consumo de drogas, bajo rendimiento académico, alto grado de ausentismo escolar, deserción, bajos índices de actividades extracurriculares y escasas aspiraciones educativas [11]. En este sentido, los centros escolares son el marco idóneo para el estudio de estos hábitos, como para poner en marcha programas de costumbres saludables y prevención de drogodependencias.

El objetivo de esta investigación es determinar a través del Análisis de Clases Latentes (ACL) conglomerados de jóvenes escolarizados en relación con sus patrones de consumo de drogas, para producir un conocimiento que brinde elementos que permitan direccionar las estrategias de prevención e intervención que actualmente se desarrollan en el país.

En la presente investigación son utilizados los resultados del estudio “*Factores que inciden en el consumo de drogas, población juvenil. Región Central de Occidente*” de Costa Rica. El estudio fue patrocinado por el Instituto Costarricense sobre Drogas, la Asociación Ramonense Pro Bienestar de la Comunidad, la Caja Costarricense de Seguro Social, el Ministerio de Educación, el Instituto de Alcoholismo y Fármaco dependencia y la Universidad de Costa Rica, durante noviembre del 2006.

La población en estudio está formada por 13.428 jóvenes perteneciente a 135 centros de enseñanza y 17 Equipos Básicos de Atención Integral de Salud (EBAIS). La muestra de 7.553 jóvenes se seleccionó utilizando un muestreo por conglomerados. Esta encuesta se diseñó principalmente para proporcionar información sobre el consumo de drogas, actividades de tiempo libre, participación de los jóvenes en actividades de la comunidad, condiciones de vida y niveles de información de los jóvenes. El módulo de interés son las variables respecto a “*Cuando fue que, por primera vez utilizó drogas*”; se tienen así 13 drogas (variables o ítemes) con respuestas binarias (0=no, 1=Sí)<sup>1</sup>. El listado de variables del cuestionario utilizado en la presente investigación es el siguiente: fumó cigarrillos ( $X_1$ ), bebió cerveza ( $X_2$ ), bebió vino ( $X_3$ ), consumió bebidas fuertes (whisky, vodka, ginebra, guaro, etc.) ( $X_4$ ), fumó marihuana ( $X_5$ ), consumió cocaína ( $X_6$ ), utilizó pastillas (estimulantes, tranquilizantes) ( $X_7$ ), uso inhalante ( $X_8$ ), utilizó alucinógenos ( $X_9$ ), uso heroína ( $X_{10}$ ), uso éxtasis ( $X_{11}$ ), consumió crack ( $X_{12}$ ) y utilizó otras drogas ( $X_{13}$ ).

## 2 Análisis de clases latentes

El Análisis de Clases Latentes es un método multivariante introducido por Henry & Lazarsfeld [10], este método permite estudiar la existencia de una (o varias) variable(s) latente(s) por medio 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 clásico, se supone que

---

<sup>1</sup>Los jóvenes respondieron a qué edad comenzaron a consumir las drogas, para aquellos que marcaron una edad se codificó como 1 y en caso contrario, los que no han utilizado drogas, como 0.

tanto las variables manifiestas como las variables latentes son categóricas. En Goodman [9] podemos observar la conexión que existe entre estos modelos y la teoría moderna de las tablas de contingencia. Los problemas de estimación e identificación han sido desarrollados, entre otros, por Anderson [3] y Mchugh [14]. Finalmente, se puede citar a distintos autores que han desarrollado esta metodología, como Agresti [1], Andersen [2], Bartholomew [4], Clogg [5], Vermut & Magidson [17], entre otros.

La relación entre las variables manifiestas debe verificar el principio de *Independencia local*. Este supone que dentro de cada categoría de la variable latente, las variables observadas son estadísticamente independientes, es decir, las variables de la tabla de contingencia son condicionalmente independientes dada una clase determinada de la variable latente. Así, toda la asociación observada entre las variables manifiestas, está medida o explicada por la variable latente.

Supongamos que se tiene una matriz conteniendo información de  $p$  variables categóricas sobre una muestra de  $n$  individuos,

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}$$

, donde cada fila de la matriz, contiene el conjunto de respuestas de un individuo dado para cada una de las variables, conocido como *respuesta* o *patrón de respuesta*. Un caso particular del Modelo de Clases Latentes se presenta cuando las variables manifiestas son binarias, es decir, solamente se tienen dos categorías de respuesta (0,1), por lo que la matriz anterior estaría conformada por ceros y unos.

Sea  $\mathbf{X}' = (X_1, \dots, X_p)$  un vector de  $p$  variables manifiestas binarias, las cuales conforman una tabla de contingencia  $p$ -dimensional. Supongamos que estas  $p$  variables, se consideran indicadoras de una variable latente  $Y$  con  $C$  categorías o clases latentes. El MCL que describe la situación anterior está dado por:

$$\pi_{\mathbf{X}}(\mathbf{x}) = \sum_{c=1}^C \pi_{\mathbf{X}, Y}(\mathbf{x}, c), \quad (1)$$

donde  $\pi_{\mathbf{X}, Y}(\mathbf{x}, c) = P(\mathbf{X} = \mathbf{x}, Y = c)$  es la probabilidad conjunta de que un individuo seleccionado aleatoriamente tenga un patrón de respuesta  $\mathbf{x}' = (x_1, x_2, \dots, x_p)$  y esté en la clase latente  $c$ . La expresión anterior implica que la población puede ser dividida en  $C$  clases latentes excluyentes

y exhaustivas, por lo tanto la probabilidad conjunta de las variables manifiestas se obtiene sumando sobre la dimensión latente. En este sentido, esta expresión implica (y justifica) la existencia de la variable latente.

Siguiendo con el modelo (1) y considerando el principio de independencia local, la probabilidad conjunta está dada por:

$$\pi_{\mathbf{X},Y}(\mathbf{x}, c) = \pi_Y(c) \prod_{i=1}^p \pi_{X_i/Y(c)}(x_i), \quad (2)$$

donde  $\pi_Y(c) = P(Y = c)$ , representa la proporción de elementos que se encuentran en la clase latente  $c$ , también conocida como probabilidad *a priori*, y  $\pi_{X_i/Y(c)}(x_i)$  representa la probabilidad condicional  $P(X_i = x_i | Y = c)$ . Ya que las variables manifiestas son binarias, estas probabilidades condicionales siguen una distribución Bernoulli:

$$\pi_{X_i/Y(c)}(x_i) = \pi_{ic}^{x_i} (1 - \pi_{ic})^{1-x_i}, \quad (3)$$

donde  $\pi_{ic}$  es la probabilidad condicional de obtener una respuesta positiva en la variable  $X_i$  para un individuo de la clase latente  $c$ .

Así, el MCL lo podemos reescribir como:

$$\pi_{\mathbf{X}}(\mathbf{x}) = \sum_{c=1}^C \pi_Y(c) \prod_{i=1}^p \pi_{ic}^{x_i} (1 - \pi_{ic})^{1-x_i}. \quad (4)$$

En la ecuación (4) queda claramente reflejado que al estimar un MCL debemos obtener los estimadores de las probabilidades *a priori*,  $\pi_Y(c)$ , y de las probabilidades condicionales,  $\pi_{ic}$ , los cuales se obtienen mediante el algoritmo EM [6], considerando la restricción

$$\sum_{c=1}^C \pi_Y(c) = 1 \quad (5)$$

y además, considerando que

$$0 \leq \pi_{ic} \leq 1. \quad (6)$$

Una vez que el modelo ha sido estimado, los individuos se clasifican dentro de la clase latente más probable, considerando las probabilidades *a posteriori*:

$$\pi_{Y/\mathbf{X}(\mathbf{x})}(c) = \frac{\pi_{\mathbf{X},Y}(\mathbf{x}, c)}{\pi_{\mathbf{X}}(\mathbf{x})} = \frac{\pi_Y(c) \pi_{\mathbf{X}/Y(c)}(\mathbf{x})}{\pi_{\mathbf{X}}(\mathbf{x})}. \quad (7)$$

En la práctica, para cada patrón de respuesta se inspecciona este conjunto de probabilidades, y se asigna el individuo a la clase latente en la cual esta probabilidad es mayor (asignación modal).

Por otra parte, la bondad de ajuste de un MCL se determina comparando las frecuencias observadas para cada patrón de respuesta,  $f_x$ , con las frecuencias estimadas,  $\hat{f}_x$ , mediante el contraste Chi-cuadrado de Pearson o la razón de verosimilitud  $G^2$ :

$$\chi^2 = \sum_x \frac{(f_x - \hat{f}_x)^2}{\hat{f}_x} \quad (8)$$

$$G^2 = 2 \sum_x f_x \ln \left( \frac{f_x}{\hat{f}_x} \right). \quad (9)$$

Ambos estadísticos siguen una distribución Chi-cuadrado con  $2^p - m - 1$  grados de libertad, donde  $m$  es el número de parámetros a estimar.

Para evaluar un MCL también se utiliza el Indice de disimilaridad, definido como

$$I_D = \sum_x \frac{|f_x - \hat{f}_x|}{2n}. \quad (10)$$

Este índice toma valores entre 0 y 1. En general, los valores del índice menores que 0.05 se consideran aceptables.

Cuando se tienen muchas celdas con frecuencias bajas, pueden surgir problemas con la distribución teórica de los estadísticos de bondad de ajuste [12]. Una alternativa a los test anteriores, es presentada por Read & Cressie [15], ellos proponen la utilización de una versión generalizada del estadístico Chi-cuadrado:

$$CR(\lambda) = \frac{2}{\lambda(\lambda+1)} \sum_x f_x \times \left[ \left( \frac{f_x}{\hat{f}_x} \right)^\lambda - 1 \right], \quad (11)$$

donde  $0 \leq \lambda \leq 1$ .

Otra solución al problema de las frecuencias bajas es la utilización del método bootstrap, desarrollado y publicado por Bradley Efron [7]. La esencia del *método bootstrap* consiste en que en ausencia de otra información, los valores de una muestra aleatoria son la mejor representación de la distribución de la población y remuestrear la muestra nos proporciona la mejor información sobre lo que sucedería si remuestreáramos la población [8] [13]. En este caso hablamos de *Bootstrap no paramétrico*.

En algunas ocasiones se tiene algún conocimiento más que el estrictamente aportado por la muestra, por ejemplo se conoce la función de distribución de la variable objeto de estudio pero se desconocen los parámetros que deben ser estimados. En estos casos se estimarían los parámetros a partir de la muestra y el remuestreo se realizaría a partir de la función teórica conocida, con los parámetros estimados, en lugar de a partir de la función de distribución empírica construida a partir de la muestra. este procedimiento se conoce como *Bootstrap Paramétrico*, el cual puede considerarse para las pruebas de bondad de ajuste en un MCL [18].

A este respecto, Von Davier [18], establece que a un nivel de significación ( $\alpha$ ), si al menos  $(n^*\alpha)^2$  muestras *bootstrap* tienen un estadístico de bondad de ajuste mayor al observado con los datos originales, el modelo es rechazado con un nivel de  $(1 - \alpha)\%$  de confianza. Langeheine et al. [12] afirman que un modelo no se ajusta a los datos, si la proporción  $\alpha$  de los  $G^2$  obtenidos con *bootstrap* que son mayores que el  $G^2$  original es pequeña.

### 3 Resultados

En el Cuadro 1 se presentan las distribuciones de frecuencias marginales y bivariadas del consumo de drogas. Tanto en filas como en columnas las drogas se identifican con números del 1 al 13 según el orden expuesto con anterioridad. En la diagonal de la matriz se presenta la distribución de frecuencias de las drogas; por ejemplo, 2.441 jóvenes tienen el hábito de fumar cigarrillos (droga 1) y 166 han consumido cocaína (droga 6).

Las frecuencias bivariadas nos revelan por ejemplo, que 2.300 personas tienen el hábito de fumar (droga 1) y beber cerveza (droga 2) o que 147 jóvenes consumen cocaína (droga 6) y marihuana (droga 5). Se observa una “asimetría” de la matriz, determinada por el consumo de marihuana, que resulta como una frontera entre las drogas débiles (o legales) y las fuertes (las prohibidas).

Por otra parte, considerando que se tiene 13 variables manifiestas binarias, el número de posibles patrones de respuesta es de 8.192; y aunque el tamaño de muestra es elevado (7.553), el 98% de los patrones de respuesta no fueron observados. En el Cuadro 2, se presentan los patrones de respuestas observados en la muestra, que equivalen aproximadamente a un 2% del total de patrones posibles. Por ejemplo, la frecuencia del patrón (0000000000000) nos indica que 1.513 personas no consumen ninguna de

---

<sup>2</sup> $n^*$  es el número de muestras *bootstrap* consideradas

		Drogas												
		1	2	3	4	5	6	7	8	9	10	11	12	13
1		<b>2.441</b>												
2		<b>2.300</b>	4.573											
3		3.239	4.202	5.588										
4		831	2.840	2.886	3.016									
5		498	576	580	554	601								
6		145	155	154	154	147	<b>166</b>							
7		125	129	131	127	109	94	139						
8		121	133	137	128	110	95	92	145					
9		98	103	102	104	98	84	87	86	110				
10		79	80	82	83	79	79	75	76	78	89			
11		82	91	91	92	85	80	82	81	81	99			
12		96	100	101	102	97	92	87	88	81	83	85	109	
13		135	145	145	144	109	89	91	92	82	80	88	92	157

Cuadro 1: Matriz de frecuencias marginales y bivariadas de factores que inciden en el consumo de drogas.

la drogas y por el contrario, 56 jóvenes consumen todas las drogas (patrón de respuesta (1111111111111)).

Patrón (x)	Frecuencia	Patrón (x)	Frecuencia
<b>00000000000000</b>	<b>1513</b>	1111101000000	6
11110000000000	1214	1111100100000	5
00100000000000	1168	0110000100000	5
01100000000000	1015	0011100000000	5
01110000000000	925	1111010000000	4
11100000000000	400	1100100000000	4
11111000000000	281	0111000100000	4
01000000000000	189	0010000000001	4
00110000000000	111	1111111111110	3
11000000000000	69	1111111110000	3
01111000000000	63	1111111100011	3
11010000000000	59	1111111100000	3
<b>111111111111</b>	<b>56</b>	1111110000010	3
10100000000000	56	1110001000000	3
10000000000000	43	1101100000000	3
01010000000000	32	1000100000000	3
11111100000000	29	0010100000000	3
10110000000000	23	1111101000101	2
11101000000000	18	1111001000001	2
00010000000000	17	1111000001111	2
1111000000001	16	1110000100000	2
1111100000001	14	1100000000001	2
01101000000000	13	0111110000000	2
11111000100000	9	0111010000000	2
1111001000000	8	0110010000000	2
1111000100000	8	0011111111111	2
10010000000000	8	0010000100000	2
1111110100000	6	Otros*	107

\* Otros patrones respuesta, todos con frecuencias igual a 1.

Cuadro 2: Distribución de los patrones de respuestas observados.

Debido a que nos encontramos con una tabla de frecuencias poco ocupada, la selección del modelo adecuado para clasificar a los jóvenes en grupos homogéneos en relación al consumo de drogas, se llevó a cabo considerando el método *bootstrap paramétrico*<sup>3</sup>, con el propósito de determinar la distribución de los estadísticos de bondad de ajuste (*EBA*). La decisión del modelo escogido está basada en la distribución *bootstrap* de los estadísticos de bondad de ajuste, utilizando los criterios propuestos por Von Davier [18] y Langeheine et al. [12].

Para encontrar el modelo adecuado, se procedió al ajuste de múltiples modelos, incrementando consecutivamente el número de clases latentes. Los modelos en que  $C = 1, \dots, 7$  no son apropiados. Ambos criterios concuerdan en que no hay una buena bondad de ajuste. Si  $C = 8$ , sólo hay tres valores de  $G^2$  y  $\chi^2$  que son mayores al calculado a partir de la muestra original (ver Cuadros 3 y 4), por lo que no rechazamos este modelo, considerando un nivel de confianza del 99%. Así, la tipología de jóvenes consumidores de drogas puede ser explicada por 8 subgrupos o clases. Además al considerar el *índice de disimilaridad* (en inglés, dissimilarity index)<sup>4</sup>, el modelo con ocho clases latentes, da lugar a una clasificación errónea de sujetos mal asignados de apenas un 2%.

Estadísticos	Valor original ( $n = 7.553$ )	Bootstrap Paramétrico	
		Media	Desviación estándar
Chi-cuadrado	66.131,62	34.042,07	178.329,17
Razón de Verosimilitud	575,55	346,11	250,46
Cressie-Read	5.602,28	2.768,98	11.110,85

Cuadro 3: Valores originales, media y desviación estándar de los EBA obtenidos con bootstrap paramétrico ( $B = 100$ , donde  $B$  es el número de remuestras bootstrap).

---

<sup>3</sup>Para realizar los cálculos de los modelos se utilizó el programa WINMIRA, versión 2001 [19]

<sup>4</sup>El índice de disimilaridad,  $I_D$ , es definido en términos de las frecuencias observadas y esperadas. Como regla general, valores de  $I_D$  menores a 0,05 son considerados pequeños.

Chi-cuadrado	Razón de Verosimilitud	Cressie-Read
216,73	149,84	171,62
429,37	191,04	265,1
985,65	160,6	355,74
690,58	193,05	362,19
...	...	...
27.709,44	359,15	2.746,11
45.643,81	289,22	2.901,27
44.496,86	313,62	3.192,02
<b>66.131,62</b>	<b>575,55</b>	<b>5.602,28</b>
574.735,26	1.820,02	49.931,25
1.135.742,61	1.557,16	55.270,69
1.292.008,87	1.676,19	86.402,94

Cuadro 4: Resultados de los EBA para las remuestras bootstrap.

### 3.1 Interpretación de las clases latentes

En el Cuadro 5 se puede apreciar la caracterización de las clases latentes, considerando las probabilidades condicionales de que un joven utilice la droga  $i$  ( $i=1,2,\dots,13$ ) dado que se encuentra en la clase latente  $c$  ( $c = 1, 2, \dots, 8$ )<sup>5</sup>.

La primera clase latente (C1) se puede interpretar como la correspondiente al grupo de jóvenes “alcohólicos”, debido a que solamente tienen el hábito de consumir bebidas alcohólicas: cerveza (0,9003), vino (0,9803) y vodka (0,8219); la probabilidad de pertenecer a esta clase es 0,0891. Estas personas constituyen el grupo más tolerado por las sociedades.

La segunda (C2) y sexta (C6) clase latente tienen probabilidades de pertenencia de 0,0171 y 0,0094 respectivamente. Están compuestas por jóvenes que consumen bebidas alcohólicas, fuman cigarrillos y marihuana, pero en la segunda clase el consumo de vodka (1,0000) y marihuana (1,0000) se encuentra más asentado. Además, en la clase sexta (C6) los jóvenes utilizan otros tipos de drogas (0,6120). En tanto, los sujetos que componen la tercera (C3) clase con una probabilidad de pertenecer de 0,3849, puede ser considerada como los jóvenes abstemios, o sea, los que no consumen ningún tipo de droga, ya que las probabilidades condicionales son pequeñas para todas las drogas. La cuarta (C4) clase (con probabili-

---

<sup>5</sup>Cada clase latente estará caracterizada por aquellas variables para las que estas probabilidades condicionales son mayores.

Drogas	Probabilidades totales de las clases							
	C1	C2	C3	C4	C5	C6	C7	C8
Cigarrillos	0,0891	0,0171	0,3849	0,2841	0,0185	0,0094	0,1870	0,0099
Cerveza	0,0008	0,9705	0,0126	0,2793	1,0000	0,8431	1,0000	0,8935
Vino	0,9003	0,9836	0,0903	0,8995	0,7038	0,9667	1,0000	0,9068
Vodka	0,9803	0,9893	0,4156	0,9563	0,2134	0,8871	0,9754	0,9336
Marihuana	0,8219	1,0000	0,0088	0,3796	0,2519	0,9274	0,9358	0,9467
Cocaína	0,1265	1,0000	0,0008	0,0000	0,0606	0,6411	0,1832	0,9605
Pastillas	0,0048	0,3599	0,0002	0,0016	0,0000	0,4758	0,0041	0,9739
Inhalantes	0,0000	0,0958	0,0006	0,0019	0,0000	0,5190	0,0081	0,9718
Alucinógenos	0,0038	0,1030	0,0009	0,0063	0,0000	0,4498	0,0059	0,9737
Heroína	0,0012	0,0925	0,0000	0,0006	0,0000	0,2822	0,0009	1,0000
Éxtasis	0,0000	0,0000	0,0003	0,0000	0,0000	0,3285	0,0000	1,0000
Crack	0,0000	0,0424	0,0000	0,0000	0,0071	0,3924	0,0000	1,0000
Otras	0,0013	0,0800	0,0018	0,0005	0,0171	0,6120	0,0163	0,9467

Cuadro 5: Parámetros del modelo clásico de ocho clases latentes: probabilidades totales y condicionadas

dad de 0,2841), son personas que solamente toman cerveza (0,8995) y vino (0,9563).

La quinta clase latente (C5) tiene una probabilidad de pertenencia de 0,0185 y está formada por los que tienen el hábito de fumar (1,0000) y tomar cerveza (0,7038). La séptima (C7) clase consumen las drogas legales principalmente cigarrillos (1,0000) y cerveza (1,0000), con menor frecuencia beben vino (0,9734) y vodka (0,9358), los jóvenes tienen una probabilidad de pertenecer a la clase de 0,1870. Los casos más graves son sujetos politoxicómanos que definen la octava (C8) clase latente y la probabilidad que un adolescente pertenezca a la clase es escasamente de 0,0099. Se observa que todas las drogas tienen probabilidades altas con relación a la clase. La politoxicomanía o policonsumo se produce cuando una persona se administra una variada gama de drogas.

### 3.2 Caracterización demográfica de las clases latentes

Hemos clasificado a los adolescentes en ocho clases latentes y con ello, se obtuvo una segmentación de los jóvenes en grupos. Trataremos ahora de describir las características demográficas de los jóvenes pertenecientes a cada uno de estos.

Con relación al sexo de los jóvenes encuestados, los varones representan la mayoría en las clases anteriormente analizadas, solamente en la clase 3 y 4 predominan las mujeres con un 58,4% y 52,4% respectivamente. En tanto, la clase ocho (o sea, los que consumen todas las drogas) está constituida por un 79,5% de hombres.

Los adolescentes que poseen mayor probabilidad de pertenecer a la clase 4 son los más pequeños con una edad media de 13,31 años, y seguidamente los que no consumen ningún tipo de droga (clase 3) con un promedio de 14,7 años. Por otra parte, los sujetos de la clase 5 que tienen el hábito de fumar y tomar cerveza son mayores de edad tal que promedian los 16,57 años. En términos generales, podemos concluir que los promedios de edad de los adolescentes que forman los ocho grupos no todas son iguales (Prob.=0,000). Utilizando el método de Bonferroni, con el propósito de ajustar el valor para cada prueba individual de igualdad de medias, se concluye que la edad promedio de la clase 4 es significativamente diferente a todas las otras clases. Entre otras diferencias, se destaca que las clases 1, 3 y 8 tienen promedios de edad significativos menores a las clases 2, 5, 6 y 7.

En todas las clases latentes se observa que los adolescentes están estudiando, pero en los politoxicómanos el 13,7% ha dejado de estudiar. En relación con el último año aprobado la mayoría tienen el colegio, solamente el 60,7% de los jóvenes de la clase 4 (vino – cerveza) son de escuela.

El porcentaje de jóvenes que han trabajado o realizaron tareas en casa en los últimos 12 meses es pequeño (menos del 7%) en los ocho grupos. Los que trabajaron y estudiaron tiende a aumentar, presentándose el porcentaje más grande en la clase 2, o sea, los que consumen las drogas legales y marihuana (38.5%). En tanto, el menor porcentaje se tiene en la clase 3 (“sanos”) con un 20.7%.

En la clase 1, 3, 4, 6 y 7 los adolescentes dedicaron los últimos 12 meses solamente a estudiar: 57.0%, 69.5%, 74.0% 53.5% y 58.9%, respectivamente; por lo contrario, en la clase 2 el 46.2% de las personas dedicaron el tiempo exclusivamente en estudiar. Es importante mencionar que los que no trabajaron y no hicieron nada en especial en los últimos 12 meses alcanzan 20.8% en la clase 8, que resulta significativamente mayor que en todos los otros grupos.

#### 4 Conclusiones

Tal y como se advierte a partir de lo previamente expuesto y de los resultados reseñados, el modelo de clases latentes nos permite llegar a las siguientes conclusiones: hay un grupo importante de jóvenes que no son consumidores de ningún tipo de droga (38%), un pequeño número de personas son adictos a todas las drogas (1%), existen un conjunto jóvenes que solamente consumen bebidas alcohólicas (9%) y la mayoría de personas utilizan drogas legales (58%).

Los rasgos característicos de las diferentes clases (o segmentos) contribuyen a comprender ciertos aspectos cualitativos del consumo y utilización de las drogas, explorar su heterogeneidad y clasificar a las personas en distintas categorías para poder caracterizarlos según variables demográficas. Así por ejemplo, los hábitos de consumo son diferentes en ambos sexos, especialmente es importante destacar que las mujeres representan la mayoría en los abstemios y los que han probado vino. Los “policonsumidores” son más frecuentes a medida que aumenta la edad de los jóvenes.

Podemos afirmar que este trabajo presenta un método alternativo para estudiar el problema del uso y utilización de drogas, al utilizar el modelo de clases latentes y así incorporar un enfoque multidimensional al análisis de los resultados, a diferencias de los trabajos que se han publicado a nivel nacional, que tienen limitaciones importantes: por lo general se limitan a estudiar el problema utilizando métodos univariados como distribuciones de frecuencias y pueden no reflejar el proceso que subyace al problema

multivariante de la drogadicción.

## Referencias

- [1] Agresti, A. (1984) *Analysis of Ordinal Categorical Data*. John Wiley and Sons, New York.
- [2] Andersen, E. (1991) *The Analysis of Categorical Data*. Springer-Verlag, Berlin.
- [3] Anderson, T. (1954) “On estimation of parameters in latent structure analysis”. *Psychometrika* **19**(1): 1–10.
- [4] Bartholomew, D. (1987) *Latent Variable Models and Factor Analysis*. 2nd edition, Oxford University Press, London.
- [5] Clogg, C.C. (1995) “Latent class models: recent developments and prospects for the future”, in: C.C. Clogg; G. Arminger & M.E. Sobel (Eds.) *Handbook of Statistical Modeling in the Social Sciences*, Plenum, New York: 311–359.
- [6] Dempster, A.P.; Laird, N.M.; Rubin, D.B. (1977) “Maximum likelihood from incomplete data via the EM algorithm”, *Journal of the Royal Statistical Society* **39**: 1–38.
- [7] Efron, B. (1979) “Bootstrap Methods: Another Look at the Jackknife”, *The Annals of Statistics* **7**: 1–26.
- [8] Efron, B.; Tibshirani, R. J. (1993) *An Introduction to the Bootstrap*. Chapman and Hall, London.
- [9] Goodman, L. (1974) “Exploratory latent structure analysis using both identifiable and unidentifiable models”, *Biometrika* **61**(2): 215–231.
- [10] Henry, N.; Lazarsfeld, P. (1968) *Latent Structure Analysis*. Houghton Mifflin, Boston.
- [11] Instituto Sobre Alcoholismo y Farmacodependencia (2007) *La Juventud y las Drogas: Encuesta Nacional sobre Percepciones y Consumo en Población de Educación Secundaria*. San José, Costa Rica.
- [12] Langeheine, R.; Pannekoek, J.; Van de Pol, F. (1996) “Bootstrapping goodness-of-fit measures in categorical data Analysis”, *Sociological Methods & Research* **24**(4): 492–516.

- [13] Manly, F. (1997) *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Chapman & Hall, London.
- [14] McHugh, R. (1956) “Efficient estimation and local identification in latent class analysis”, *Psychometrika* **21**(4): 331–347.
- [15] Read, T.; Cressie, N. (1988) *Goodness-of-Fit Statistics for Discrete Multivariate Data*. Springer-Verlag, New York.
- [16] Ugalde, M.; Barboza, A.; Orozco, J.; Esquivel, J.; Jiménez, J.; Gómez, A. (2000) “Análisis del consumo de drogas en adolescentes escolarizados de Costa Rica”, *Revista de la Asociación de Ciencias Penales de Costa Rica* **18**(24): 79–84.
- [17] Vermunt, J.; Magidson, J. (2002) “Latent class cluster analysis”, in: J.A. Hagenaars & A.L. McCutcheon (Eds.) *Applied Latent Class analysis*. Cambridge University Press, London: 89–106.
- [18] Von Davier, M. (1997) “Bootstrapping goodness-of-fit statistics for sparse Categorical Data – Results of a Monte Carlo study”, *Methods of Psychological Research Online* **2**: 29–48.
- [19] Von Davier, M. (2001) *WINMIRA 3.2 pro*. Institute for Science Education (IPN), Kiel.