

## UNA CARACTERIZACIÓN DE LA SATISFACCIÓN ESTUDIANTIL MEDIANTE ANÁLISIS DE CLASES LATENTES

RAMÓN ÁLVAREZ-VAZ<sup>a</sup>, ELENA VERNAZZA<sup>a</sup>

<sup>a</sup>Instituto de Estadística, Departamento de Métodos Matemáticos Cuantitativos, Facultad de Ciencias Económicas y de Administración, Universidad de la República, Uruguay.  
e-mail: ramon@iesta.edu.uy, evernazza@iesta.edu.uy

En este trabajo se estudian las principales características de la construcción de la Satisfacción Estudiantil en los cursos de grado de la Facultad de Ciencias Económicas y de Administración, Universidad de la República, Uruguay, mediante Análisis de Clases Latentes (ACL). Los datos utilizados para la aplicación presentada en este trabajo provienen de una encuesta aplicada sobre una muestra de estudiantes de grado de la Facultad, en el año 2009. Dicho cuestionario presenta una estructura de bloques: por un lado (primer bloque) se encuentran las variables que permitirán realizar una caracterización sociodemográfica de los estudiantes y por otro se presentan las variables del modelo ECSI (European Customer Satisfaction Index) que serán las utilizadas para la caracterización de la Satisfacción Estudiantil. Las variables manifiestas consideradas como insumo para la construcción y caracterización de la Satisfacción Estudiantil son las siguientes 6: expectativas (E) de los estudiantes al ingresar al centro de estudios, la imagen (I) que tienen de éste, la calidad de la enseñanza recibida (CSA) y de los servicios brindados (CSF), las necesidades y deseos personales con respecto a la facultad (ND) y el valor percibido (VP). Los resultados presentados surgen de considerar que efectivamente existe una variable que refiere a la Satisfacción Estudiantil y que ésta queda definida, a partir de la interacción de las 6 variables manifiestas, por 4 clases latentes.

**Keywords:** Caracterización, Clases Latentes, Estudiantes Universitarios, Satisfacción.

### 1. Introducción

Conocer el nivel de satisfacción de los clientes, con un determinado servicio que se les brinda resulta fundamental como insumo en la toma de decisiones que tengan como objetivo primordial mantener o mejorar, en caso de que sea necesario, aquellos aspectos que se entiende determinan la *Satisfacción*.

Vinculando esta idea con la educación universitaria, se toma lo propuesto por Alves y Raposo [6], quienes plantean: “Sólo con la satisfacción de los alumnos se podrá alcanzar el éxito escolar, la permanencia de los estudiantes en la institución y, sobre todo, la formación de una valoración positiva boca a boca. En este sentido, es extremadamente importante encontrar formas fiables de medir la satisfacción del alumno en la enseñanza universitaria, permitiendo así a las instituciones de enseñanza conocer su realidad, compararla con la de los otros competidores y analizarla a lo largo del tiempo”.

En este trabajo se considera a los estudiantes universitarios que concurren a la Facultad de Ciencias

Económicas y Administración (FCEA) de la Universidad de la República, como “clientes” y se determina que el “servicio” que se les brinda es el de la educación de nivel terciario.

La información necesaria para poder establecer cómo se construye el concepto de *Satisfacción*, se obtiene a través de la aplicación de un cuestionario formado por apartados de preguntas que conforman el modelo ECSI (European Customer Satisfaction Index). Sobre este instrumento, y a través del Análisis de Clases Latentes, se analiza cómo se construye la *Satisfacción Estudiantil* [21], [5].

La estructura de este documento es la siguiente: en primera instancia se hace referencia a las características generales de la técnica estadística empleada. A continuación se presenta una descripción de los datos, seguida de los principales resultados obtenidos y por último, se plantean las principales conclusiones y propuestas de líneas de trabajo e investigaciones a futuro.

## 2. Metodología

Generalmente, al trabajar con datos categóricos multivariados, resulta de interés investigar eventuales fuentes de confusión entre las variables observadas, identificar/caracterizar grupos de individuos y aproximar la distribución de las observaciones a través de las variables en estudio [15]. Existe una técnica que contempla todas estas situaciones: Análisis de Clases Latentes (ACL) o Modelos de Clases Latentes (MCL) [14], [11], [7], [12], [1].

El ACL busca segmentar la tabla/hipercubo de contingencia creado a partir de las variables observadas/*manifiestas*, por una variable no observada/*latente*, con la siguiente característica: se supone que las respuestas a todas las variables manifiestas son estadísticamente independientes con respecto a los valores de la variable de clases latentes (*independencia local o condicional*). De este manera el modelo asocia, en términos de probabilidad, a cada individuo a una *clase* latente. Se puede predeterminar, por lo tanto, el valor esperado con el que una observación responde a cada variable manifiesta. Si bien el modelo estimado no estipula el número de *clases* latentes, pueden usarse varios estadísticos de bondad de ajuste para evaluar de forma tanto teórica como empírica la cantidad de clases a considerar.

Este método puede verse como un modelo de regresión y, por lo tanto, sería posible incluir variables predictivas para la membresía de cada observación a una clase latente [7].

Existen antecedentes de estudios con este tipo de variables en disciplinas como la economía y la psicología. En particular, en el trabajo “Segmentación de la población española según su grado de concienciación ecológica mediante modelos de variables latentes”, los autores presentan una segmentación de los hábitos de consumo en función de su grado de concienciación ecológica mediante técnicas estadísticas de ACL [19].

Por otra parte, en el trabajo “Modelos De Clases Latentes Para Definir Perfiles Conductuales en Niños De 4 y 5 Años” [9] sus autores elaboran perfiles conductuales en niños de 4 o 5 años de México aplicando ACL sobre los resultados del test de screening “Child Behavior Check List” (CBCL).

En otras áreas de la salud como la epidemiología, se presenta el trabajo “Análisis de clases latentes en tablas poco ocupadas: consumo de alcohol, tabaco y otras drogas en adolescentes” [8], en el que sus autores presentan una segmentación del tipo de consumo de drogas en jóvenes de Costa Rica.

**2.1. Definición del Modelo.** Se considera un modelo en el que se observan  $J$  variables categóricas politómicas (variables *manifiestas*) tal que cada una tiene  $K_j$  posibles respuestas, para los  $i = 1, 2, \dots, N$  individuos.

El modelo de clases latentes aproxima la distribución conjunta observada de las variables manifiestas como la suma ponderada (por un número finito  $R$ ) de las tablas de clasificación cruzada.

$Y_{ijk}$  será el valor observado de las  $J$  variables manifiestas para el individuo  $i$ , tal que  $Y_{ijk} = 1$  si el individuo  $i$  da la respuesta  $k$  de la variable  $j$  y  $Y_{ijk} = 0$  en otro caso, con  $j = 1, \dots, J$  y  $k = 1, \dots, K_j$  y  $\pi_{jrk}$  representará la probabilidad condicional de que una observación en la clase  $r = 1, \dots, R$  produzca el  $k$ -ésimo resultado de la variable  $j$ -ésima.

Dentro de cada clase, para cada variable manifiesta, se cumple:

$$\sum_k \pi_{jrk} = 1.$$

Por otra parte,  $p_r$  corresponderá a las proporciones a partir de las cuales serán generados los pesos para la suma ponderada de las tablas de clasificación ( $\sum_r p_r = 1$ ). En este sentido, considerando que estos  $p_r$  representan la probabilidad *incondicional* de que un individuo pertenezca a una clase (antes de tomar en cuenta el valor de  $Y_{ijk}$ ),  $p_r$  será denominado probabilidad *a priori* de la membresía a cada clase latente.

La probabilidad de que un individuo  $i$  en la clase  $r$  genere un conjunto  $J$  de resultados en las variables manifiestas, asumiendo independencia condicional de los resultados  $Y$  dado la pertenencia a una clase dada, es:

$$f(Y_i; \pi_r) = \prod_{j=1}^J \prod_{k=1}^{K_j} (\pi_{jrk})^{Y_{ijk}} \quad (1)$$

Además, la función de densidad es:

$$P(Y_i | \pi, p) = \sum_{r=1}^R p_r \prod_{j=1}^J \prod_{k=1}^{K_j} (\pi_{jrk})^{Y_{ijk}} \quad (2)$$

De esta manera se tienen 2 parámetros a estimar por el modelo:  $p_r$  y  $\pi_{jrk}$ . Dadas  $\hat{p}_r$  y  $\hat{\pi}_{jrk}$ , las probabilidades *a posteriori* de que cada individuo pertenezca a una clase latente, condicionada a los valores observados de las variables manifiestas, queda determinada:

$$\hat{P}(r_i | Y_i) = \frac{\hat{p}_r f(Y_i; \hat{\pi}_r)}{\sum_{q=1}^R \hat{p}_q f(Y_i; \hat{\pi}_q)} \quad (3)$$

con  $r_i = 1, \dots, R$ .

Se debe tener en cuenta que las  $\hat{\pi}_{jrk}$  son estimaciones de las probabilidades de los resultados condicionales en la clase  $r$ . También es importante observar que el número de parámetros independientes estimados aumenta rápidamente con  $R$ ,  $J$  y  $K_j$ . Dados estos valores, el número de parámetros a estimar es  $R \sum_j (K_j - 1) + (R - 1)$ . Este último resultado puede producir una situación no deseada, ya que cuando este número excede el número total de observaciones, o una menos que el número total de celdas en la tabla de clasificación cruzada de las variables manifiestas, el modelo no puede ser identificado.

**2.2. Estimación de parámetros.** Los modelos de clase latente pueden estimarse mediante máxima verosimilitud, donde la log-verosimilitud es:

$$\ln L = \sum_1^N \ln \sum_1^R p_r \prod_{j=1}^J \prod_{k=1}^{K_j} (\pi_{jrk})^{Y_{ijk}} \quad (4)$$

Dicha verosimilitud será maximizada con respecto a  $p_r$  y  $\pi_{jrk}$  a través del algoritmo *EM* [10], [16]. Como con cualquier modelo de mezcla finita, el algoritmo *EM* se puede aplicar en virtud de que la membresía a la clase de cada individuo es desconocida, por lo que se trata como un problema de datos faltantes. El algoritmo trabaja en forma iterativa en 2 fases a partir de valores iniciales arbitrarios de  $\hat{p}_r$  y  $\hat{\pi}_{jrk}$ , los que se etiquetan como  $\hat{p}_r^{anterior}$  y  $\hat{\pi}_{jrk}^{anterior}$ .

1. En la fase de esperanza (E), se calcula la probabilidad de membresía a la clase *latente* usando la ecuación (3), sustituyendo en los valores  $\hat{p}_r^{anterior}$  y  $\hat{\pi}_{jrk}^{anterior}$ .
2. En la fase de maximización (M) los parámetros estimados se actualizan maximizando la log-verosimilitud dada la probabilidad a *posteriori* ( $\hat{P}(r_i | Y_i)$ ).

La nueva probabilidad a *priori*, será:

$$\hat{p}_r^{nueva} = \frac{1}{N} \sum_{i=1}^N \hat{P}(r_i | Y_i) \quad (5)$$

y la nueva probabilidad condicional será:

$$\hat{\pi}_{jrk}^{nueva} = \frac{\sum_{i=1}^N Y_{ijk} \hat{P}(r_i | Y_i)}{\sum_{i=1}^N \hat{P}(r_i | Y_i)} \quad (6)$$

En la ecuación (6),  $\hat{\pi}_{jrk}^{nueva}$  es el vector de longitud  $K_j$  de las probabilidades condicionales para la  $j$ -ésima variable manifiesta; y por otra parte  $Y_{ij}$  es la matriz  $N \times K_j$  de resultados para  $Y_{ijk}$  para esa variable. Como todo proceso iterativo este algoritmo repite las 2 fases sustituyendo el valor viejo por el nuevo, hasta

alcanzar un máximo o hasta que el incremento que tiene la log-verosimilitud sea menor a un cierto umbral previamente establecido.

Por último, cabe destacar que la aplicación de este método de estimación depende de: los valores iniciales elegidos para  $\hat{p}_r^{anterior}$  y  $\hat{\pi}_{jrk}^{anterior}$  y de la complejidad del modelo que se estima, por lo que el algoritmo *EM* puede encontrar un máximo local de la función log-verosimilitud, en lugar del máximo global deseado, con lo cual es recomendable estimar más de una vez.

### 2.3. Criterios de selección y validación del modelo.

Tal como fuera mencionado previamente, la estimación a través de ACL no estipula una cantidad de *clases* latentes, sin embargo una de las ventajas de esta técnica, a diferencia de varias de las técnicas de clusterización más comúnmente utilizadas, es la variedad de herramientas existentes para determinar dicha cantidad.

En algunos de los casos, el número de clases latentes a utilizar viene dado por la teoría o por experiencias empíricas previas. Sin embargo, en la mayoría de los casos será necesario realizar un análisis exploratorio que permita decidir la cantidad de clases latentes presentes en el problema en estudio. Este proceso comienza presentando el modelo más general posible, es decir, un modelo con independencia completa que determina una sola clase. Una vez estimado dicho modelo, el número de clases se va incrementando de una en una hasta encontrar el modelo que resulte el “modelo adecuado”.

Agregar una clase al modelo mejorará el ajuste, pero incorporará ruido y parámetros a estimar, por lo que será necesario tener en consideración un criterio de parsimonia que establezca un equilibrio entre la mejora del ajuste y la cantidad de parámetros que se incorporan al aumentar una clase en el modelo. El criterio de parsimonia utilizado en este trabajo será el del mínimo BIC (Bayesian information criterion).

## 3. Resultados

En esta sección se presentan, en forma resumida, los datos con los que se trabajó, describiendo el diseño muestral empleado y el cuestionario utilizado. Por último se exponen los principales resultados del Análisis de Clases Latentes.

**3.1. Diseño muestral.** La aplicación que se presentará en este trabajo fue realizada sobre los datos obtenidos mediante la aplicación de un cuestionario sobre una muestra de estudiantes universitarios de la FCEA, en el año 2009.

La muestra fue seleccionada en base a un marco muestral que se construyó a partir de las inscripciones a cursos de FCEA en dicho año. El diseño muestral usado fue estratificado por conglomerados en dos etapas y la muestra finalmente quedó conformada por estudiantes de 60 grupos prácticos (repartidos en forma proporcional en 6 estratos, uno por cada año y otro específico con asignaturas de la Licenciatura en Administración). Con la muestra seleccionada, se realizó el relevamiento de datos que culminó con 647 encuestas realizadas (tasa de cobertura de la muestra de 90 %).

**3.2. Cuestionario utilizado.** El cuestionario, aplicado sobre la muestra seleccionada, resulta de una adaptación del cuestionario utilizado por los investigadores Alves y Raposo de la Universidad de Beira Interior (Portugal) [6]. Éste presenta la siguiente estructura: un primer bloque, claramente diferenciado de los demás, que contiene una serie de variables de carácter sociodemográfico, como sexo, edad y algunas otras variables que caracterizan al estudiante dentro del ámbito de la facultad, como año de ingreso, año y cantidad de materias en curso, entre otras. Los restantes ocho bloques de preguntas (presentados como bloque A - H) presentan todos la misma estructura, se plantea una pregunta general que determina la esencia del bloque y a partir de ella, se establecen una serie de afirmaciones sobre las cuales el estudiante deberá expresar su posición, utilizando una escala Likert que toma valores en el intervalo [1 - 10], donde 1 indicará la mayor discrepancia con lo planteado en la pregunta y 10 el mayor acuerdo.

Los bloques A a H presentan las siguientes características:

- Bloque A - 12 afirmaciones referentes a las *expectativas* de los estudiantes, previo ingreso a facultad.
- Bloque B - 6 afirmaciones vinculadas a la *imagen* que tienen los estudiantes sobre la facultad.
- Bloque C - 9 afirmaciones asociadas a la *calidad* del servicio que brinda la facultad.
- Bloque D - 9 afirmaciones asociadas a la *calidad* de los servicios que brinda la facultad con respecto a la biblioteca, bedelía y cafetería, entre otros.
- Bloque E - 9 afirmaciones (las mismas que el bloque C) asociadas a necesidades/deseos actuales.
- Bloque F - 7 afirmaciones que indagan sobre el *valor percibido*.
- Bloque G - 6 afirmaciones que refieren a la *satisfacción* de los estudiantes con la facultad.
- Bloque H - 5 afirmaciones que pueden dividirse en 2 subgrupos, las 3 primeras referentes a la *lealtad*

de los estudiantes con la facultad, y las 2 últimas asociadas al *boca a boca* que se genera entre los estudiantes.

Las variables consideradas en este trabajo resultan de la agregación de los ítems por Bloques: Expectativas ( $E_R$ ), Imagen ( $I_R$ ), Calidad de Servicios Académicos ( $CSA_R$ ), Calidad de Servicios Funcionales ( $CSF_R$ ), Necesidades/Deseos ( $ND_R$ ) y Valor Percibido ( $VP_R$ ).

**3.3. Aplicación.** La aplicación presentada en este trabajo, toma como insumo las puntuaciones (categorizadas) de las seis variables del ECSI.

A modo de simplificar la notación, la codificación utilizada será (para  $i = 1:6$ ):

$$Y_i = \begin{cases} 3 & \text{Alto} \\ 2 & \text{Medio} \\ 1 & \text{Bajo} \end{cases}$$

$$Y = (E_R, I_R, CSA_R, CSF_R, ND_R, VP_R)$$

En resumen, se tiene:

- Tamaño de muestra:  $n = 470$  (luego de depurar los datos y descartar observaciones con datos faltantes).
- Una **variable de clases latentes**: Satisfacción estudiantil.
- Seis variables ( $Y_i$ ) **manifiestas**:  $p = 6$ .
- Cada una de las variables manifiestas posee 3 categorías de respuestas posibles:  $k_i = 3$  (para  $i = 1:6$ ).

En la Tabla 1 se presentan los seis patrones de respuesta más frecuentes (de las 163 secuencias observadas), y sus respectivas frecuencias, para el caso de los 470 estudiantes en estudio. En dicha tabla se puede observar que los dos patrones más frecuentes son los que representan los extremos: niveles *altos de todas* las variables manifiestas y valores *bajos en todas* ellas, respectivamente.

Table 1. Patrones de respuesta - Frecuencias observadas.

$E_R$	$I_R$	$CSA_R$	$CSF_R$	$ND_R$	$VP_R$	Frecuencia
3	3	3	3	3	3	48
1	1	1	1	1	1	45
2	2	2	3	2	2	17
1	2	1	1	1	1	11
3	3	3	3	3	2	11
3	2	3	3	3	2	10

**3.4. Estimación del modelo.** Los modelos estimados, presentados en esta sección fueron estimados con el paquete *poLCA* [15] del Software libre R-project [17].

En el contexto del análisis de variables de clases latentes estimar un modelo consiste, en primera instancia, en determinar cuántas clases latentes existen en el problema en estudio.

Por lo tanto, la hipótesis de partida en la estimación de cada uno de los posibles modelos será:

$H_0$  : El modelo ajustado es el *adecuado*.

$H_1$  : El modelo ajustado NO es el *adecuado*.

Por *adecuado* se entenderá que la cantidad de clases especificadas es la correcta.

En este trabajo se han estimado 4 modelos ( $M = 1, 2, 3, 4$ ) y a partir de los resultados obtenidos se determina que, siguiendo el criterio de mínimo BIC, el mejor modelo sería aquel que presenta una variable con 3 clases latentes. Sin embargo, tanto en este modelo como en aquellos que proponen una variable con una y dos clases latentes los resultados obtenidos ponen de manifiesto que la hipótesis nula es rechazada<sup>1</sup>, por lo que sería necesario un ajuste con más clases.

Para el caso del modelo con una variable con cuatro clases latentes la hipótesis nula no puede ser rechazada, por lo que podría considerarse que ajustar un modelo con cuatro clases latentes es *adecuado*. Además, se verifica que al estimar este modelo no existen problemas de identificabilidad y que en el proceso de maximización se alcanza, al menos, un máximo local (que puede coincidir con el máximo global).

**3.5. Caracterización de las clases.** En función de lo expuesto se decide estimar un modelo de una variable con cuatro clases latentes, cuyas probabilidades a priori son:  $P(1) = 0.32$ ,  $P(2) = 0.29$ ;  $P(3) = 0.24$  ;  $P(4) = 0.15$ .

La caracterización de cada una de las clases se realiza en función de la probabilidad condicional, de cada una de las categorías de cada variable manifiesta, dada la clase. Tomando como referencia los resultados presentados en la Tabla 2, la caracterización de las clases en las que se agrupan a los 470 estudiantes es la siguiente:

#### Clase Latente 1

Los estudiantes que se encuentran en esta clase presentan un nivel de *expectativas* y una percepción de la *calidad de los servicios funcionales* medio-bajo y niveles medios

de *imagen*, percepción de la *calidad de los servicios académicos*, *necesidades y deseos* y *valor percibido*.

#### Clase Latente 2

Los estudiantes que se encuentran en la clase latente 2, presentan un nivel alto de todas las variables manifiestas. Cabe destacar, además, el hecho de que la probabilidad de que un estudiante que pertenece a esta clase, tenga niveles bajos en su percepción de la *calidad de los servicios académicos*, es 0.

#### Clase Latente 3

En el extremo opuesto a los estudiantes cuya *Satisfacción* se define a partir de la clase latente 2, se encuentran los estudiantes de esta clase. Éstos presentan un nivel bajo de todas las variables manifiestas. Cabe destacar, además, el hecho de que la probabilidad de que un estudiante que pertenece a esta clase, tenga niveles altos en *imagen* y *necesidades/deseos*, es 0.

#### Clase Latente 4

Por último, los estudiantes que pertenecen a la clase latente 4 se caracterizan por tener nivel medio-alto de *expectativas*, nivel medio de *imagen*, percepción de la *calidad de los servicios académicos*, *necesidades/deseos* y *valor percibido*. En lo que refiere a la percepción de la *calidad de los servicios funcionales*, los estudiantes que se encuentran en esta clase presentan valores altos. Además, se destaca que la probabilidad de que un estudiante que pertenece a esta clase, tenga niveles bajos en *imagen*, percepción de la *calidad los servicios académicos* y *necesidades/deseos*, es 0.

Por lo tanto, en lo que refiere a la *Satisfacción Estudiantil*, las clases latentes se podrían categorizar como:

- $m = 1$ : Estudiantes con *Satisfacción Estudiantil* **medio-baja**.
- $m = 2$ : Estudiantes con *Satisfacción Estudiantil* **alta**.
- $m = 3$ : Estudiantes con *Satisfacción Estudiantil* **baja**.
- $m = 4$ : Estudiantes con *Satisfacción Estudiantil* **media-alta**.

<sup>1</sup>Se considera un  $\alpha = 0,05$ .



Table 2. Probabilidades condicionales  $P(Y_i/m)$ .

$E_R$	1	2	3
m = 1	0.42	0.43	0.15
m = 2	0.04	0.15	0.81
m = 3	0.81	0.17	0.02
m = 4	0.09	0.48	0.43
$I_R$	1	2	3
m = 1	0.19	0.62	0.19
m = 2	0.01	0.26	0.73
m = 3	0.70	0.30	0.00
m = 4	0.00	0.91	0.09
$CSA_R$	1	2	3
m = 1	0.30	0.69	0.01
m = 2	0.00	0.05	0.95
m = 3	0.99	0.00	0.01
m = 4	0.00	0.68	0.32
$CSF_R$	1	2	3
m = 1	0.38	0.40	0.22
m = 2	0.04	0.16	0.80
m = 3	0.77	0.19	0.04
m = 4	0.03	0.29	0.68
$ND_R$	1	2	3
m = 1	0.28	0.63	0.09
m = 2	0.01	0.07	0.92
m = 3	0.97	0.03	0.00
m = 4	0.00	0.74	0.26
$VP_R$	1	2	3
m = 1	0.13	0.57	0.30
m = 2	0.03	0.24	0.73
m = 3	0.75	0.24	0.01
m = 4	0.05	0.95	0.00

**3.6. Probabilidades a posteriori.** Los resultados presentados en la Tabla 3 hacen referencia, a modo de ejemplo, a las probabilidades a posteriori, para cada uno de los patrones posibles de respuesta (para los 6 patrones más frecuentes, presentados en la Tabla 1), y la asignación a cada una de las clases (en función de la máxima probabilidad a posteriori).

Table 3. Probs. a posteriori según patrones y asignaciones.

$E_R$	$I_R$	$CSA_R$	$CSF_R$	$ND_R$	$VP_R$	$P(1)$	$P(2)$	$P(3)$	$P(4)$	Asignación
3	3	3	3	3	3	0.00	1.00	0.00	0.00	2
1	1	1	1	1	1	0.00	0.00	1.00	0.00	3
2	2	2	3	2	2	0.18	0.00	0.00	0.81	4
1	2	1	1	1	1	0.01	0.00	0.99	0.00	3
3	3	3	3	3	2	0.00	0.99	0.00	0.01	2
3	2	3	3	3	2	0.00	0.76	0.00	0.24	2

En función de las probabilidades a posteriori, de cada una de las clases, el total de estudiantes queda distribuido

en cada una de ellas en un 30 % (145), 29 % (137), 25 % (118) y 16 % (77) respectivamente.

#### 4. Conclusiones

Los resultados obtenidos y presentados en este trabajo, nos permiten concluir:

- Efectivamente existe una variable de clases latentes que refiere a la *Satisfacción Estudiantil* de los estudiantes de nivel universitario de la Facultad de Ciencias Económicas y Administración (FCEA, UDELAR, Uruguay).
- Dicha variable tiene 4 clases que quedan definidas a partir de la interacción de las 6 variables manifiestas.
- En función de la pertenencia a cada una de las clases, se detectan estudiantes universitarios cuya satisfacción con la FCEA es extrema en 2 sentidos opuestos: *Satisfacción Estudiantil* alta y *Satisfacción Estudiantil* baja. En un término medio, existen estudiantes cuya *Satisfacción Estudiantil* se define como media. Para estos, existen dos clases: media-alta y media-baja.

#### 5. Consideraciones finales

Como consideraciones finales y propuestas a futuro se plantea:

- Evaluar la robustez de la variable de clases latentes, iterando varias veces para evaluar el grado de dependencia de los valores iniciales.
- Plantear el uso de las clases latentes detectadas para, a través de una partición de la tabla de datos, volver a estimar un Modelo de Ecuaciones Estructurales y realizar un estudio comparativo con los modelos ya estimados con estos datos [21], [2], [3], [4].
- Estudiar la creación de variables de clases latentes a través del uso de *Mixture models: latent profile and latent class analysis* [18] o [20], trabajando sobre las variables manifiestas en su escala original, es decir, las 6 variables sin categorizar.
- Evaluar la asociación entre las categorías de las variables manifiestas a través de *graph model* (construyendo un grafo entre las categorías observadas) y mediante el análisis de redes (SNA) [13].
- Realizar un estudio similar con los datos obtenidos en investigación realizada en 2017 (réplica del estudio 2009). Comparar resultados.

## Referencias

- [1] Agresti, A. (2013). *Categorical data analysis*. Wiley-Interscience, Hoboken, N.J.
- [2] Álvarez-Vaz, R. y Vernazza, E. (2013). Aplicación de los modelos de ecuaciones estructurales para el estudio de la satisfacción estudiantil en los cursos superiores de fceeya. *Documentos de Trabajo - IESTA. Serie DT (13 / 02)*.
- [3] Álvarez-Vaz, R. y Vernazza, E. (2014). Aplicación de modelos de ecuaciones estructurales en la medición del nivel de satisfacción estudiantil: comparación de tres métodos de estimación. *Documentos de Trabajo - IESTA. Serie DT (14 / 03)*.
- [4] Álvarez-Vaz, R. y Vernazza, E. (2017). Evaluación de un instrumento de medición del nivel de satisfacción estudiantil a través de la aplicación de modelos de ecuaciones estructurales. *Cuadernos del CIMBAGE*, (19):1–25.
- [5] Álvarez-Vaz, R., Freira, D., Vernazza, E., y Alves, H. (2016). Can students' satisfaction indexes be applied the same way in different countries? *Int Rev Public Nonprofit Marketing*, 13(101).
- [6] Alves, H. y Raposo, M. (2004). La medición de la satisfacción en la enseñanza universitaria: El ejemplo de la universidade da beira interior. *Int Rev Public Nonprofit Marketing*, 1(1):73–88.
- [7] Bandeen-Roche, K., Miglioretti, D. L., Zeger, S. L., y Rathouz, P. J. (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association*, 92(440):1375–1386.
- [8] Carlomagno, A. A. y Sepúlveda, R. (2010). Análisis de clases latentes en tablas poco ocupadas: consumo de alcohol, tabaco y otras drogas en adolescentes. *Revista de Matemática: Teoría y Aplicaciones*, 17(1):25–40.
- [9] Castro López, C. R., Montano Rivas, A., y Oliva Zarate, L. (2011). Modelos de clases latentes para definir perfiles conductuales en niños de 4 y 5 años. *Revista Electrónica de Psicología Iztacala*, 14(1).
- [10] 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*, 39:1–38.
- [11] Everitt, B. S. (1984). *An Introduction to Latent Variable Models*. Springer Netherlands, Dordrecht.
- [12] Hagenaars, J. (2002). *Applied latent class analysis*. Cambridge University Press, Cambridge New York.
- [13] Højsgaard, S. (2012). *Graphical models with R*. Springer, New York.
- [14] Lazarsfeld, P. (1950). The logical and mathematical foundations of latent structure analysis. *ISA Stouffer (ed.), Measurement and Prediction*, pp. 362–412.
- [15] Linzer, D. A. y Lewis, J. B. (2011). poLCA: An R package for polytomous variable latent class analysis. *Journal of Statistical Software*, 42(10):1–29.
- [16] McLachlan, G. (2000). *Finite mixture models*. Wiley, New York.
- [17] R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- [18] Robertson, J. (2016). *Modern statistical methods for HCI*. Springer, Switzerland.
- [19] Sánchez Rivero, M. (2001). Segmentación de la población española según su grado de concienciación ecológica mediante modelos de variables latentes I. *Investigaciones Europeas de Dirección y Economía de la Empresa*, 7(3):173–196.
- [20] Scrucca, L. y Raftery, A. E. (2015). Improved initialisation of model-based clustering using Gaussian hierarchical partitions. *ArXiv e-prints*.
- [21] Vernazza, E. (2013). Evaluación de un instrumento de medición del nivel de satisfacción estudiantil en los cursos de formación superior de la FCCEEyA de la UDELAR a través de la aplicación de Structural Equation Modelling (SEM). Informe de pasantía, Fac. de CCEEyA - UdelAR.