

# CUADERNOS DEL CIMBAGE



Universidad de Buenos Aires  
Facultad de Ciencias Económicas



---

## CARACTERIZACIÓN DE LA SATISFACCIÓN ESTUDIANTIL EN LA FACULTAD DE CIENCIAS ECONÓMICAS Y DE ADMINISTRACIÓN EN URUGUAY: UNA APLICACIÓN DE ANÁLISIS DE CLASES LATENTES Y DE ANÁLISIS DE CLUSTERS

Autor(es): Álvarez-Vaz, Ramón y Vernazza-Mañan, Elena

Fuente: Cuadernos del CIMBAGE, Nº 21 (Junio, 2019), pp. 65-85

Publicado por: Facultad de Ciencias Económicas de la Universidad de Buenos Aires

Vínculo: <http://ojs.econ.uba.ar/ojs/index.php/CIMBAGE/issue/view/198>

---



Esta revista está protegida bajo una licencia Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Copia de la licencia: <https://creativecommons.org/licenses/by-nc-sa/4.0/>



Cuadernos del CIMBAGE es una revista académica semestral editada por el Centro de Investigaciones en Metodologías Básicas y Aplicadas a la Gestión (CIMBAGE) perteneciente al Instituto de Investigaciones en Administración, Contabilidad y Métodos Cuantitativos para la Gestión (IADCOM).

## **CARACTERIZACIÓN DE LA SATISFACCIÓN ESTUDIANTIL EN LA FACULTAD DE CIENCIAS ECONÓMICAS Y DE ADMINISTRACIÓN EN URUGUAY: UNA APLICACIÓN DE ANÁLISIS DE CLASES LATENTES Y DE ANÁLISIS DE CLUSTERS**

Ramón Álvarez-Vaz\*, Elena Vernazza-Mañan\*\*

Instituto de Estadística. Facultad de Ciencias Económicas y de Administración. Universidad de la República  
Eduardo Acevedo 1139, CP.11200 Montevideo, Uruguay

\*ramon@iesta.edu.uy; \*\*evernazza@iesta.edu.uy

Recibido 23 de septiembre de 2018, aceptado 25 de marzo 2019

---

### **Resumen**

En este trabajo se presentan los principales resultados obtenidos al realizar un estudio de la caracterización de la Satisfacción Estudiantil, en los cursos de grado de la Facultad de Ciencias Económicas y de Administración, Universidad de la República, Uruguay, a través de la utilización y comparación de dos técnicas de análisis de datos multivariantes: Análisis de Clases Latentes y Análisis de Clusters. Los datos utilizados provienen de una encuesta aplicada sobre una muestra de estudiantes de la Facultad. El cuestionario presenta una estructura de 2 bloques. El primero, presenta las variables que permitirán realizar una caracterización sociodemográfica de los estudiantes, y el segundo contiene 63 ítems subdivididos en 8 escalas asociadas al modelo ECSI que serán las utilizadas para la caracterización de la Satisfacción Estudiantil. Las variables manifiestas consideradas para la construcción y caracterización de la Satisfacción Estudiantil son las siguientes: *expectativas* de los estudiantes al ingresar al centro de estudios, la *imagen* que tienen de éste, la *calidad de la enseñanza* recibida y de los *servicios brindados*, las *necesidades y deseos* personales con respecto a la Facultad y el *valor percibido*. Estas variables surgen de agrupar los ítems del ECSI, por escala. 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 cuatro clases latentes. Por otra parte, se propone agrupar a los estudiantes en tres *clusters* (utilizando el algoritmo de Ward).

**Palabras clave:** caracterización, clases latentes, clusters, satisfacción

**Códigos JEL:** C1, C3, I2

---

**CHARACTERIZATION OF STUDENT'S SATISFACTION AT  
FACULTAD DE CIENCIAS ECONÓMICAS Y DE  
ADMINISTRACIÓN, DE LA UNIVERSIDAD DE LA REPÚBLICA,  
URUGUAY: AN APLICATION OF LATENT CLASSES ANALYSIS  
AND CLUSTER ANALYSIS**

Ramón Álvarez-Vaz\*, Elena Vernazza-Mañan\*\*

Instituto de Estadística. Facultad de Ciencias Económicas y de  
Administración. Universidad de la República.  
Eduardo Acevedo 1139. CP.11200, Montevideo, Uruguay

\*ramon@iesta.edu.uy; \*\*evernazza@iesta.edu.uy

Received September 23<sup>th</sup> 2018, accepted March 25<sup>th</sup> 2019

---

**Abstract**

This article presents the main results obtained by characterizing the construction of students' satisfaction at Facultad de Ciencias Económicas y de Administración, Universidad de la República, Uruguay, through the use and comparison of two multivariate data analysis techniques: latent classes analysis and cluster analysis. The used data arise survey applied to a sample of undergraduate students of the Faculty, in the year 2009. This survey, present a structure in blocks: on the one hand, the variables that allow making a sociodemographic characterization of students. On the other hand (second block) there is the ECSI model (European Customer Satisfaction Index), which will be used to students' satisfaction characterization. The ECSI's variables are grouped in: expectations of the incoming students, the image that students have about the college, teaching and services quality, the needs and personal desires about college, and the perceived value. The main results presented in this work consider, on the one hand, that there is indeed a variable that refers to students' satisfaction and that it is defined by four latent classes, from the interaction of the 6 manifest variables. On the other hand, from the analysis of clustering through the Ward method, it is proposed to group the students into three clusters. Finally, the results' comparison obtained with both techniques it is also presented.

**Keywords:** characterization, latent classes, clusters, satisfaction

**JEL Codes:** C1, C3, I2

---

## 1. INTRODUCCIÓN

Quien brinda cualquier tipo de servicio, entiende que conocer el nivel de satisfacción de los clientes, con dicho servicio resulta fundamental como insumo en la toma de decisiones que tengan como objetivo central mantener o mejorar (si es necesario) aquellos aspectos que determinan y caracterizan la *Satisfacción*.

A partir de esta idea, y vinculándola con la educación universitaria, se toma lo propuesto por Alves y Raposo (Alves y Raposo, 2004), 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 de Administración (FCEA) de la Universidad de la República, Uruguay, 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 bloques de preguntas que reproducen el modelo ECSI (European Customer Satisfaction Index).

Sobre este instrumento, y a través del Análisis de Clases Latentes y del Análisis de Clusters, se analiza cómo se construye la *Satisfacción Estudiantil* y cómo se agrupan los estudiantes en función de las variables consideradas (Vernazza, 2013), (Álvarez-Vaz *et al.*, 2016).

El presente trabajo se estructura de la siguiente manera: en primera instancia se presenta la metodología utilizada, haciendo referencia a las principales características de las técnicas empleadas (sección 2).

La sección 3 comienza con una descripción de los datos, se presenta el diseño muestral y las variables utilizadas. A continuación, se exponen los principales resultados obtenidos a través de un Análisis de Clases Latentes y de un Análisis de Cluster (sección 4 y sección 5).

Dichos resultados son presentados, en forma conjunta y comparada, en la sección 6. Por último, en la sección 7, se exponen las principales conclusiones.

## 2. METODOLOGÍA

En esta sección se presentan las principales características de las dos técnicas estadísticas utilizadas: Análisis de Clases Latentes y Análisis de Clusters. En lo que refiere a la primera técnica, se expone brevemente su principal objetivo y se plantean sus diferentes enfoques estadísticos. Por último, se define formalmente del modelo, su forma de estimación y de validación. En cuanto al Análisis de Clusters se presenta la técnica haciendo especial énfasis en una estrategia de agrupación mediante algoritmos jerárquicos.

### 2.1 ANÁLISIS DE CLASES LATENTES

Generalmente, al trabajar con datos categóricos multivariados, resulta de interés investigar posibles fuentes de confusión entre las variables observadas, identificar y caracterizar grupos de individuos y aproximar la distribución de las observaciones a través de las variables en estudio (Linzer y Lewis, 2011). Existe una técnica que contempla todas estas situaciones: el Análisis de Clases Latentes (ACL) o Modelos de Clases Latentes (MCL) (Lazarsfeld, 1950), (Everitt, 1984), (Bandein-Roche et al., 1997), (Hagenaars, 2002), (Agresti, 2013).

El ACL busca segmentar la tabla (hipercubo) de contingencia, creada a partir de las variables observadas/*manifiestas*, por una variable no observada/*latente*, suponiendo que las respuestas a todas las variables manifiestas son estadísticamente independientes con respecto a los valores de la variable latente (independencia local/condicional). De esta 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 dicho individuo responde a cada variable observada. Si bien el modelo estimado no estipula el número de *clases* latentes presentes en el problema en estudio, existen varios estadísticos de bondad de ajuste que permiten evaluar, tanto teórica como empíricamente, la cantidad de clases a considerar.

#### 2.1.1 Definición del modelo

Sean  $J$  variables categóricas politómicas (variables manifiestas) tal que cada una tiene  $K_j$  posibles respuestas, para los  $i=1,2,\dots,N$  individuos.

$Y_{ijk}$  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:

$$\sum_k^{K_j} \pi_{jrk} = 1 \quad (1)$$

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). 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 probabilidades *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 en  $Y$  dado la pertenencia a una clase dada, es:

$$f(Y_i; \pi_r) = \prod_{j=1}^J \prod_{k=1}^{K_j} (\pi_{jrk})^{Y_{ijk}}. \quad (2)$$

Además, la función de densidad es:

$$P(Y_i | \pi, p) = \sum_{r=1}^R p_r \prod_{j=1}^J \prod_{k=1}^{K_j} (\pi_{jrk})^{Y_{ijk}} \quad (3)$$

Dadas  $\hat{p}_r$  y  $\hat{\pi}_{jrk}$ , la probabilidad *a posteriori* de que cada individuo pertenezca a una clase latente será:

$$\hat{P}(r_i | Y_i) = \frac{\hat{p}_r f(Y_i; \hat{\pi}_r)}{\sum_{q=1}^R \hat{p}_q f(Y_i; \hat{\pi}_q)} \quad (4)$$

El método de estimación utilizado en este trabajo es el de máxima verosimilitud, a través del algoritmo de maximización EM.

- En la fase de esperanza (E), se calcula la probabilidad de membresía a la clase latente.
- En la fase de maximización (M) los parámetros estimados se actualizan maximizando la log-verosimilitud dada la probabilidad *a posteriori*.

La estimación no estipula una cantidad de clases latentes. Esta se puede determinar a partir de la teoría o por experiencias empíricas previas. En la mayoría de los casos será necesario tener en consideración un criterio de parsimonia (en este caso mínimo BIC).

## 2.2 ANÁLISIS DE CLUSTERS

Existen variados algoritmos de clusterización que se agrupan, generalmente, en 2 categorías: Jerárquicos y No jerárquicos. Sobre ellos se pueden aplicar distintos tipos de distancias, tomando en cuenta métricas diferentes en función del tipo de variable considerada

(Maechler *et al.*, 2016). En este trabajo será utilizado un método de clusterización Jerárquico.

### **2.2.1 Métodos Jerárquicos – Método de Ward**

Los métodos jerárquicos se caracterizan por generar una serie de particiones encajadas. Al comienzo, cada objeto es asignado a su propio grupo y se inicia un proceso iterativo: en cada etapa se unen los dos grupos “más similares” y se continúa hasta que solo quede un grupo. En cada etapa las distancias definidas entre las agrupaciones se recalculan y se actualizan en función del método de agrupación que se esté utilizando.

En este trabajo será utilizada un algoritmo de agrupación jerárquica, en particular, el método de Ward. Este consiste en descomponer la variación total en: variación dentro de los grupos (*within*) y variación entre los grupos (*between*). Al estar frente a una partición dada, el método unirá aquellos grupos que produzcan el efecto de hacer mínima la variación *within* en la nueva partición.

En formato matricial:

$$T = W + B \quad (5)$$

donde  $T$  es la matriz de varianzas y covarianzas del total,  $W$  la matriz de varianzas y covarianzas dentro de los grupos y  $B$  la matriz de varianzas y covarianzas entre grupos.

## **3. APLICACIÓN**

En esta sección se presenta en forma resumida los datos con los que se trabajó, describiendo el diseño muestral empleado y el cuestionario utilizado.

### **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 de los cursos superiores 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 2009. El diseño muestral usado fue estratificado por conglomerados en dos etapas y presentó las siguientes características: en una primera instancia se formaron seis estratos (en base a una clasificación desarrollada en conjunto por investigadores de la cátedra de Metodología de Investigación y del Instituto de Estadística, FCEA,

Universidad de la República (IESTA)) que corresponden aproximadamente a cada uno de los cinco años en los en los que podía estar cada estudiante en el 2009. Adicionalmente, se propone un 6to estrato para un grupo reducido de materias que corresponden únicamente a la Licenciatura en Administración. Una vez conformados los estratos, se determina que la muestra total se repartirá en forma proporcional a la matrícula de cada estrato.

Al tener definidas las unidades de muestreo, se procede a seleccionar la muestra, proceso que presentó las siguientes etapas:

1. Se sortean los grupos prácticos de cada materia en cada estrato con probabilidad proporcional a la matrícula de cada grupo (conglomerado)
2. Mediante muestreo aleatorio simple (MAS), se seleccionan la misma cantidad de estudiantes en cada grupo seleccionado en la primera etapa. La cantidad de estudiantes de cada grupo es la misma en los seis estratos.

La muestra finalmente queda conformada por estudiantes que provienen de 60 grupos prácticos (repartidos en forma proporcional en los 6 estratos). Se sortean 12 estudiantes por grupo, lo que determina un tamaño de muestra de 720 estudiantes.

Con la muestra seleccionada, se procedió a realizar el relevamiento de los datos, el cual culminó con 647 encuestas realizadas. Esto determina una tasa de cobertura de la muestra de 90%.

### **3.2 CUESTIONARIO**

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) (Alves y Raposo, 2004). 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 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

considerados en este trabajo son los que refieren a los siguientes constructos: *expectativas* (12 ítems), *imagen* (6 ítems), *calidad de los servicios académicos* (9 ítems), *calidad de los servicios funcionales* (9 ítems), *necesidades/deseos* (9 ítems) y *valor percibido* (7 ítems).

#### 4. RESULTADOS – ANÁLISIS DE CLASES LATENTES

La aplicación presentada en este trabajo, toma como insumo las puntuaciones (categorizadas) de las seis variables del ECSI (bloques A – F).

Estas variables, y sus respectivas categorías, se presentan a continuación:

- **E: Expectativas**  
Altas (> 90), Medias (81:90), Bajas (< 81)
- **I: Imagen**  
Alta (> 50), Media (41:50), Baja (< 41)
- **CSA: Calidad de los Servicios Académicos**  
Alta (> 70), Media (61:70), Baja (< 61)
- **CSF: Calidad de los Servicios Funcionales**  
Alta (> 65), Media (56:65), Baja (< 56)
- **ND: Necesidades/Deseos**  
Alta (> 70), Media (61:70), Baja (< 61)
- **VP: Valor Percibido**  
Alta (> 60), Media (51:60), Baja (< 51)

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

$$Y_i = \begin{matrix} 3 & \textit{Alto} \\ 2 & \textit{Medio} \\ 1 & \textit{Bajo} \end{matrix}$$

Se tiene, entonces:

- Tamaño de muestra:  $n = 470$  (considerando únicamente datos completos por individuo).
- Una **variable de clases latentes**: Satisfacción estudiantil.
- Seis variables  $Y_i$  **manifiestas**:  $J = 6$ .

- Cada una de las variables manifiestas posee 3 categorías de respuestas posibles:  $k_i = 3$  (para  $i = 1:6$ ).

En la Tabla N° 1 se presentan los seis patrones de respuesta más frecuentes (de los 163 patrones observados), y sus respectivas frecuencias, para el caso 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.

E	I	CSA	CSF	ND	VP	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

Tabla N° 1. Patrones de respuesta – Frecuencias observadas.

#### 4.1 ESTIMACIÓN DEL MODELO

Los modelos estimados, presentados en esta sección fueron estimados con el paquete polCA (Linzer y Lewis, 2011) del Software libre R-project (R Core Team, 2017).

En el contexto del análisis de variables 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:

H0) El modelo ajustado es el adecuado.

H1) El modelo ajustado NO es el adecuado.

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

A continuación (Tabla N° 2), se presenta un resumen de los principales resultados obtenidos al estimar dichos modelos.

A partir de los resultados presentados en la Tabla N° 2 se observa 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 (al 5%), por lo que sería necesario un ajuste con más clases.

Clases	BIC	Chi2	valor-p	G2	valor-p
M = 1	6259.27	11755.38	0	1963.49	0
M = 2	5220.88	1408.42	0	844.92	0
M = 3	4964.77	804.16	0	508.66	0.01
M = 4	4992.73	750.84	0	456.4	0.15

Tabla N° 2. Estimación de 4 modelos – M = 1,2,3,4.

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 identificación y que en el proceso de maximización se alcanza, al menos, un máximo local (que podría coincidir con el máximo global).

#### 4.2 CARACTERIZACIÓN DE LAS CLASES

En función de lo expuesto en la sección 4.1 se decide estimar un modelo de una variable con cuatro clases latentes, cuya caracterización se presenta a continuación. Tal como se observa en la Tabla N° 3, la probabilidad de pertenecer a la clase 1 es la mayor, mientras que la menor corresponde a la clase 4.

Clase	1	2	3	4
P(m)	0.32	0.29	0.24	0.15

Tabla N° 3. Probabilidad de cada una de las clases.

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 N° 4, 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*.

En función de la descripción realizada se entiende que los patrones característicos de esta clase son:

(E, I, CSA, CSF, ND, VP) = (1, 2, 2, 1, 2, 2)

(E, I, CSA, CSF, ND, VP) = (2, 2, 2, 2, 2, 2)

Además, en cuanto a la distribución de los estudiantes de esta clase, según las variables de caracterización, se tiene:

- Más del 60% de los estudiantes son mujeres.
- Más del 50% de los estudiantes tienen entre 21 y 27 años de edad.
- Más del 55% de los estudiantes se encuentran cursando su segundo y tercer año en FCEA.

### *Clase Latente 2*

Los estudiantes que se encuentran en la clase latente 2, presentan un nivel alto de todas las variables manifiestas. 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.

El patrón específico de esta clase es:

(E, I, CSA, CSF, ND, VP) = (3, 3, 3, 3, 3, 3)

Por otra parte, en lo que refiere a las variables características de los estudiantes, en esta clase, se tiene:

- Un 57% de los estudiantes son mujeres.
- Más del 55% de los estudiantes se encuentran en el primer grupo de edad, es decir, tienen menos de 21 años de edad.
- Casi el 50% de los estudiantes se encuentran cursando su primer año en FCEA.

### *Clase Latente 3*

En el extremo opuesto a los estudiantes cuya Satisfacción se define a partir de la clase 2, se encuentran los estudiantes de esta clase. Estos

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.

El patrón específico de esta clase es:

(E, I, CSA, CSF, ND, VP) = (1, 1, 1, 1, 1, 1)

Además, la distribución de los estudiantes de esta clase, según las variables de caracterización, indica que:

- Casi un 55% de los estudiantes son hombres.
- Más del 40% de los estudiantes tiene más de 24 años de edad.
- Casi un 45% de los estudiantes se encuentran cursando su cuarto y quinto año en FCEA.

#### *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 de los servicios académicos* y *necesidades/deseos*, es 0.

Los patrones característicos de esta clase son:

(E, I, CSA, CSF, ND, VP) = (2, 2, 2, 3, 2, 2)

(E, I, CSA, CSF, ND, VP) = (3, 2, 2, 3, 2, 2)

En lo que respecta a las variables características de los estudiantes, en esta clase, se tiene:

- Más de la mitad de los estudiantes son mujeres.
- Más del 75% de los estudiantes son menores de 24 años.
- Más del 50% de los estudiantes se encuentran cursando los dos primeros años en FCEA.

<b>E</b>	<b>1</b>	<b>2</b>	<b>3</b>
<b>m = 1</b>	0.42	0.43	0.15
<b>m = 2</b>	0.04	0.15	0.81
<b>m = 3</b>	0.81	0.17	0.02
<b>m = 4</b>	0.09	0.48	0.43
<b>I</b>	<b>1</b>	<b>2</b>	<b>3</b>
<b>m = 1</b>	0.19	0.62	0.19
<b>m = 2</b>	0.01	0.26	0.73
<b>m = 3</b>	0.70	0.30	0.00
<b>m = 4</b>	0.00	0.91	0.09
<b>CSA</b>	<b>1</b>	<b>2</b>	<b>3</b>
<b>m = 1</b>	0.30	0.69	0.01
<b>m = 2</b>	0.00	0.05	0.95
<b>m = 3</b>	0.99	0.00	0.01
<b>m = 4</b>	0.00	0.68	0.32
<b>CSF</b>	<b>1</b>	<b>2</b>	<b>3</b>
<b>m = 1</b>	0.38	0.40	0.22
<b>m = 2</b>	0.04	0.16	0.80
<b>m = 3</b>	0.77	0.19	0.04
<b>m = 4</b>	0.03	0.29	0.68
<b>ND</b>	<b>1</b>	<b>2</b>	<b>3</b>
<b>m = 1</b>	0.28	0.63	0.09
<b>m = 2</b>	0.01	0.07	0.92
<b>m = 3</b>	0.97	0.03	0.00
<b>m = 4</b>	0.00	0.74	0.26
<b>VP</b>	<b>1</b>	<b>2</b>	<b>3</b>
<b>m = 1</b>	0.13	0.57	0.30
<b>m = 2</b>	0.03	0.24	0.73
<b>m = 3</b>	0.75	0.24	0.01
<b>m = 4</b>	0.05	0.95	0.00

Tabla N° 4. Probabilidades condicionales – P(Yi/m).

Por lo tanto, en lo que refiere a la Satisfacción Estudiantil, las clases 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**.

#### 4.3 PROBABILIDADES A POSTERIORI

Los resultados presentados en la Tabla N° 5 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 N° 2), y la asignación a cada una de las clases (en función de la máxima probabilidad a posteriori).

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.

E	I	CSA	CSF	ND	VP	P(m = 1)	P(m = 2)	P(m = 3)	P(m = 4)	Asig.
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

Tabla N° 5. Probabilidades a posteriori según patrones y asignaciones.

#### 5. RESULTADOS – ANÁLISIS DE CLUSTERS

En lo que refiere a los principales resultados obtenidos de la realización de una clusterización a partir de un método Jerárquico (en particular, Ward) y considerando distancias euclídeas, lo primero que se evalúa es la cantidad de clusters a determinar.

Gráficamente, en la Figura 1, se observa que resultaría adecuado considerar una estructura de 3 o 5 clusters.

Si se consideran, además, los indicadores  $R^2$ , del *pseudo-F* y *pseudo-t<sup>2</sup>* se observa, que el mayor incremento en el valor del  $R^2$  se da al pasar de considerar 2 a 3 grupos, lo que reafirmaría una posible estructura de 3 grupos (ver Tabla N° 6).

Por otra parte, se tiene que si bien el máximo valor del *pseudo-F* se da al considerar una estructura de 2 grupos, el decremento de dicho estadístico al considerar 3 grupos (11% menos), no resulta tan amplio como en los demás pasos (por ejemplo, al pasar de 3 a 4 la caída es de más de un 20%).

Al observar el *pseudo-t<sup>2</sup>*, se nota una caída muy pronunciada (79%) al pasar de considerar 2 a 3 clusters por lo que la variabilidad se minimiza al considerar una potencial estructura de 3 grupos. En 5 también hay una caída, pero la de 3 resulta más fuerte.

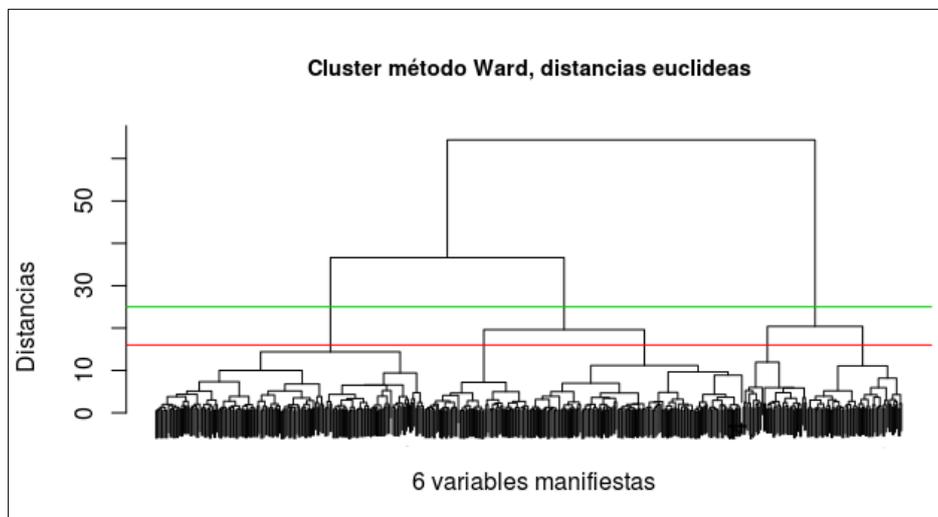


Figura N° 1: Dendrograma al usar el algoritmo de Ward para 6 variables manifiestas

<b>Paso</b>	<b>Historia</b>	<b>Frecuencia</b>	<b>R<sup>2</sup></b>	<b>ps-F</b>	<b>ps-t<sup>2</sup></b>
467 - 446	465	62	0.74	149.86	23.43
468 - 463	460	108	0.73	157.69	30.33
469 - 451	464	61	0.71	166.10	27.45
470 - 461	467	136	0.70	182.14	26.35
471 - 457	455	41	0.68	201.16	22.09
472 - 468	466	172	0.65	223.95	41.07
473 - 462	470	203	0.61	249.66	78.64
474 - 471	469	102	0.57	317.33	39.15
475 - 472	473	375	0.43	357.74	187.81
476 - 475	474	477	0.00	-	357.74

Tabla N° 6. Estadísticos - Últimos 10 pasos

Así, teniendo en consideración tanto la información gráfica como el análisis numérico de los indicadores, se decide trabajar con 3 clusters.

### 5.1 CARACTERIZACIÓN DE LOS GRUPOS

La distribución de estudiantes, por cluster, queda dada por: 36%, 43%, 31% en el cluster 1, 2 y 3, respectivamente.

#### *Cluster 1*

Los estudiantes que se encuentran en este cluster se caracterizan por ser un 54% mujeres y un 46% hombres. En lo que refiere a la edad de los estudiantes y al año que están cursando, se observa que casi un 30% se encuentra cursando su primer año en facultad, lo que resulta en concordancia con el hecho de que casi un 40% se encuentra en el primer grupo de edad (18-20). En particular, la distribución conjunta de esas dos categorías es de casi un 25% del total.

En lo que refiere a las 6 variables manifiestas, en la figura 2 se observa que el 75% de los estudiantes tienen *expectativas* más bajas que el 25% de los estudiantes del cluster 2. Además, el 25% de los estudiantes, tienen *expectativas* más altas que el 75% de los estudiantes del cluster 3. Un comportamiento similar se da para las restantes 5 variables manifiestas. Por último, al observar la figura 3, se destaca en primer lugar que existe una correlación positiva entre todas las variables y que la distribución conjunta 2 a 2 de las 6 variables manifiestas para el caso del cluster 1 presenta valores medios, en todos los casos.

### *Cluster 2*

Los estudiantes que se encuentran en el cluster 2 se caracterizan por ser más de un 60% de mujeres. En cuanto a la edad de los estudiantes y al año que están cursando, se observa que más de un 65% se encuentra cursando sus primeros años en facultad (Año en

curso = 1 y 2), lo que, tal como ocurre en el cluster 1, resulta en concordancia con el hecho de que más de un 75% se encuentra en los primeros grupos de edad (18-23). Por último, cabe destacar que en este cluster no existen estudiantes mayores de 27 años, que estén cursando su primer año en facultad.

Al analizar las variables manifiestas para el caso particular del cluster 2, se observa que el máximo de todas ellas se da en este cluster. En particular, se destaca que el 100% de los estudiantes presenta valores más altos que un 25% de los estudiantes del cluster 3, en las variables *necesidades/deseos* y *calidad de los servicios académicos* (ver Figuras N° 2 y N° 3).

### *Cluster 3*

El cluster 3 es el que presenta mayor proporción de hombres (52%). Además, casi un 60% de los estudiantes de éste está cursando su tercer, cuarto o quinto año en FCEA.

En cuanto a las variables del modelo ECSI, se observa en la figura 2 que el mínimo de todas ellas se da en este grupo. Además, el 75% de los estudiantes de este cluster, presentan valores más bajos que el 25% de los estudiantes del cluster 1, en todas las variables.

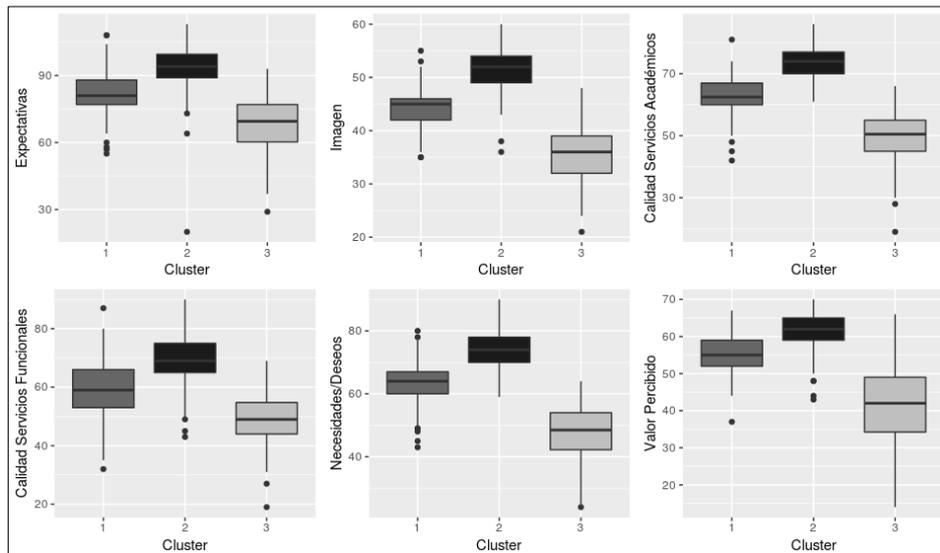


Figura N° 2: Distribución de las 6 variables manifiestas según cluster.

En función de las caracterizaciones presentadas, se entiende que en cuanto a la satisfacción estudiantil los clusters podrían denominarse:

- cluster = 1: Estudiantes con Satisfacción Estudiantil media.
- cluster = 2: Estudiantes con Satisfacción Estudiantil alta.
- cluster = 3: Estudiantes con Satisfacción Estudiantil baja.

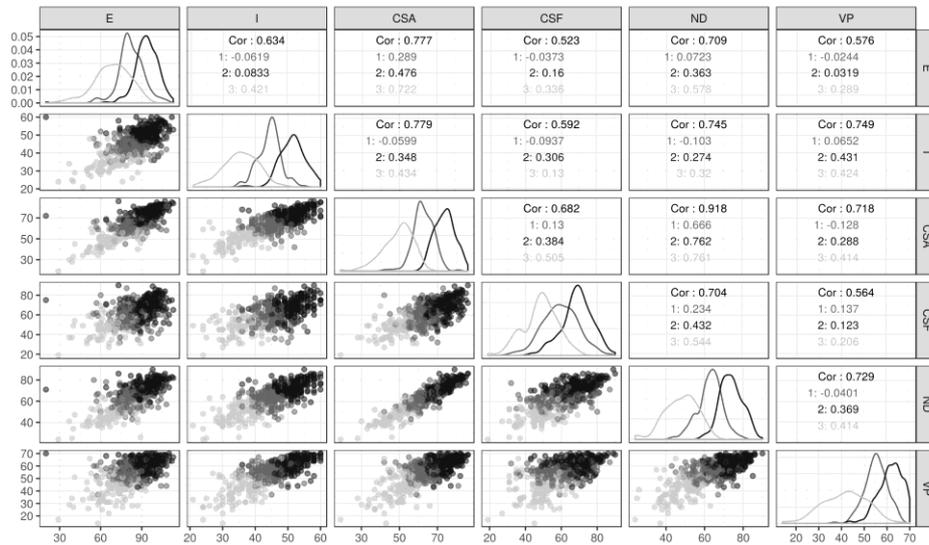


Figura 3: Diagrama de dispersión variables manifiestas.

### 6. DISCUSIÓN

Al realizar un análisis comparativo entre los resultados obtenidos con ambas técnicas, se obtiene la información resumida en la Tabla N° 7.

Cluster/Clase latente	1	2	3	4	Total
1	106	3	23	40	172
2	32	134	0	37	203
3	7	0	95	0	102
<b>Total</b>	<b>145</b>	<b>137</b>	<b>118</b>	<b>77</b>	<b>477</b>

Tabla N° 7. Estadísticos - Últimos 10 pasos

Al analizar la distribución conjunta de ambas categorizaciones se destaca, en primera instancia, que la categoría *alta* de ambas variables prácticamente coincide (alta: cluster = 2 y clase latente = 2). En particular, hay sólo un 2% de estudiantes de la clase latente alta (2), que no corresponden al cluster denominado alta (2). Cabe destacar, además, que no existen estudiantes cuya categoría de clase latente sea alta (2) y pertenezca al cluster baja (3). En lo que refiere a las categorías baja se observa que un 80% de los estudiantes de la clase latente baja (3) pertenecen al cluster baja (3). Además, se destaca el

hecho de que no existen estudiantes de la clase latente baja (3) que pertenezcan al cluster alta (2). Por último, se observa que el 85% de los estudiantes del cluster media corresponden a estudiantes cuya clase latente es media (media baja + media alta).

En resumen, se destaca que si bien las técnicas multivariantes utilizadas toman como insumo variables de distinta naturaleza, los resultados no presentan diferencias relevantes en la construcción/caracterización de la satisfacción estudiantil.

## 7. 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*.
- Dicha variable tiene 4 clases latentes que quedan definidas a partir de la interacción de 6 variables manifiestas de naturaleza categórica.
- 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.
- Teniendo en consideración las variables en su escala original, los estudiantes de la Facultad de Ciencias Económicas y de Administración pueden ser agrupados en 3 grupos heterogéneos entre sí y homogéneos a la interna, en lo que a construcción de la *Satisfacción Estudiantil* refiere.
- En función de la pertenencia a cada uno de los clusters, se identifican estudiantes cuya *Satisfacción Estudiantil* es: alta, media o baja.
- Al comparar los resultados que surgen al utilizar técnicas multivariantes que toman como insumo variables de distinta naturaleza, se observa que éstos no presentan diferencias relevantes en cuanto a la construcción/caracterización de la *Satisfacción Estudiantil*.

**REFERENCIAS BIBLIOGRÁFICAS**

- Agresti, A. (2013). *Categorical data analysis*. Wiley-Interscience, Hoboken, N.J.
- 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).
- 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.
- 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.
- Everitt, B. S. (1984). *An Introduction to Latent Variable Models*. Springer Netherlands, Dordrecht.
- Hagenaars, J. (2002). *Applied latent class analysis*. Cambridge University Press, Cambridge New York.
- Lazarsfeld, P. (1950). The logical and mathematical foundations of latent structure analysis. ISA Stouffer (ed.), *Measurement and Prediction*, pp. 362-412.
- 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.
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., y Hornik, K. (2016). *cluster: Cluster Analysis Basics and Extensions*. R package version 2.0.4 - For new features, see the 'Changelog' file (in the package source).
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- 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, Licenciatura en Estadística - Facultad de Ciencias Económicas y de Administración - Universidad de la República.