

# Determinación y Comparación de Perfiles Sociales y Culturales de estudiantes universitarios

Herrera Myriam<sup>1</sup>, Mallea Adriana<sup>2</sup>, Ruiz Ana María<sup>2</sup>, Millán Flavia<sup>1</sup>,

(1)Departamento de Informática UNSJ

(2)Departamento de Matemática de UNSJ

Ignacio de la Roza y Meglioli. Complejo CUIM. UNSJ  
Tel 0264-4265101

<sup>1</sup> mherrera@iinfo.unsj.edu.ar

## Resumen

Es de gran interés conocer las características sociales y culturales de los estudiantes de las carreras de informática de la Facultad de Ciencias Exactas de la UNSJ debido a que este conocimiento es fundamental, tanto para la gestión académica como para las investigaciones que refieren a la Universidad como objeto de estudio. Se hace necesario caracterizar el perfil social y cultural no sólo de los estudiantes ingresantes, sino de los que permanecen y egresan de las carreras. Esto implica el estudio del comportamiento de indicadores diferenciales.

Proponemos la determinación de tipologías a través de herramientas que se encuadran en el enfoque del Análisis Multidimensional de Datos (AMD).

**Palabras clave:** Variables-Multidimensional- distancias- segmentación.

## Contexto

Se presenta el estado de avance del proyecto de investigación “Determinación y Comparación de perfiles sociales y culturales de estudiantes universitarios ” de la FCEF N de la UNSJ (Argentina). El proyecto, aprobado por evaluación externa y financiado por la UNSJ, se encuentra inserto en el marco de las líneas de investigación de Estadística del Departamento de Informática.

Además se encuentra vinculado a cátedras de las carreras de Licenciatura en Ciencias de la Computación y Licenciatura en Sistemas de Información, de la FCEF N de la Universidad Nacional de San Juan.

## Introducción

Es de gran interés conocer las características sociales y culturales de los estudiantes universitarios debido a que este conocimiento es fundamental tanto para la gestión académica como para las investigaciones que refieren a la Universidad como objeto de estudio.

El debate se agudiza especialmente cuando se discuten las condiciones de acceso y equidad en el ingreso a los estudios superiores. La década de los 90 representó para la Argentina, entre otros países, el inicio de una serie de reformas políticas y económicas que afectaron la esencia de la organización del Estado. Surge el Estado como garante de la calidad de los productos educativos y generador de datos que deberían ser suficientes y apropiados para la toma de decisiones. Sin embargo sin un tratamiento estadístico adecuado y actualizado estos datos no se transforman en información viable para este fin

En el marco de estas reformas el Sistema de Educación Superior es impactado, alterando las relaciones entre Universidad-Estado. La generación de reformas se caracterizó por cambios en los modelos de financiamiento, exigencia de eficiencia a través de la implantación de sistemas evaluativos y presiones por relaciones estrechas con el sector productivo (García Guadilla, 2004).

En este contexto se hace necesario caracterizar el perfil social y cultural no sólo de los estudiantes ingresantes, sino de los que permanecen y egresan de la universidad. Esto implica el estudio del comportamiento de indicadores diferenciales.

Tradicionalmente, para realizar caracterizaciones socioeconómicas de poblaciones como en el caso de estudios de rendimiento académico se recurre a las técnicas de la estadística tradicional con la utilización de variables continuas (Ver por ejemplo Di Grescia, Porto, Ripani, 2002).

Consideramos que estas técnicas, al proponerse desde una óptica mayormente determinística reducen la riqueza de la información. En este proyecto proponemos la determinación de tipologías a través de herramientas que se encuadran en el enfoque del Análisis Multidimensional de Datos. Entre estas además de un análisis exploratorio tradicional se trabajará con un análisis simbólico de datos (Symbolic Data Analysis) y para el análisis confirmatorio el modelado estocástico con la aplicación de modelos loglineales y modelos probabilísticos gráficos..

### Líneas de investigación y desarrollo

El Análisis Multidimensional de Datos (AMD) en la versión de la escuela francesa, surge en la década de los 70, planteando fines menos deterministas que los de la Estadística tradicional. Su objetivo general es la búsqueda de una estructura presente en los datos, en un contexto de tipo más inductivo que deductivo, que revaloriza el rol del individuo. Su naturaleza, fundamentalmente descriptiva y el acercamiento geométrico asignan un rol muy importante a las representaciones gráficas, sobre todo en una etapa exploratoria.

En el campo de las Ciencias Sociales, este enfoque se revela como la opción ideal para el procesamiento de la información que, en la generalidad de los casos, es rica en categorías y no en continuos, de naturaleza ambigua, con grandes dificultades de diseño.

Los algoritmos desarrollados en el contexto del análisis multidimensional de datos se adaptan a diferentes niveles de complejidad de la información: datos numéricos, textuales, simbólicos. Es decir que el dato puede ser algo más que un único valor numérico resultado de la asignación de una medida o código a una unidad de análisis: puede ser una palabra, un conocimiento, una posibilidad, una conjunción de valores.

Una novedosa e interesante área de estudio se abre con los desarrollos en Análisis de Datos Simbólicos (Diday, 1997). Este método parte de una pregunta: ¿Por qué no se aprovechan en el procesamiento y análisis mismo los valiosos conocimientos de los expertos? La respuesta de la estadística clásica era que no se podían cuantificar. Se plantea en la actualidad el desafío de representarlos por expresiones a la vez simbólicas y numéricas, saber manipularlos y utilizar estas expresiones a los fines de ayudar a decidir, de mejorar el análisis, de sintetizar y de organizar nuestra experiencia y nuestras observaciones respetando más acabadamente su complejidad.

Estas técnicas valorizan, sobre todo, el poder de la clasificación como operación interpretadora, tratando de superar con nuevos algoritmos los problemas de descripción de las clases, en especial para los individuos que se encuentren en los bordes de las mismas.

Los conceptos de intención y extensión de una idea, aplicados a una clase o grupo, son fundamentales para la comprensión del objetivo del Análisis de Datos Simbólicos. Así la intención de una idea se refiere a los atributos que ella contiene y que no pueden ser suprimidos sin destruirla; la extensión de una idea son los sujetos o elementos a los cuales ella se aplica.

En el Análisis Simbólico en lugar de trabajar sobre las extensiones, es decir sobre los individuos, se reemplazan los individuos por las intenciones, aprovechando el conocimiento de los expertos.

Para el análisis categórico de datos, muy presentes en el ámbito de las Ciencias Sociales y de Educación, se utilizan en forma complementaria dos metodologías estadísticas multivariadas: El Análisis de Correspondencias Múltiples (ACM) y el Análisis de Modelos Log-Lineales (MLL). Ambos métodos suponen un toque de ingenio e intuición en la lógica y tratamiento de dichas variables además de provenir de dos tradiciones de análisis estadístico distintas —la anglosajona para los MLL y la escuela francesa para el AC—. Durante años han seguido procesos diferentes, dados sus orígenes y su diferente concepción. Con todo sólo recientemente se ha tratado de ver sus posibilidades complementarias. En este proyecto estos métodos se proponen como herramientas complementarias para el análisis de la estructura de dependencia entre múltiples variables categorizadas que incluso pueden presentar alguna jerarquía. La propuesta metodológica se aplicará a datos provenientes del SIU Guarani, SIU Kolla y encuestas diseñadas para la determinación y comparación de tipologías de estudiantes de dos facultades de la UNSJ, con perfiles diferenciados

## Resultados y Objetivos

Se espera caracterizar desde un enfoque social y cultural a los estudiantes universitarios del Departamento de Informática de la Facultad de Ciencias Exactas, Físicas y Naturales de la UNSJ, a través del análisis multidimensional de datos. Para lo cual se generará y suministrará información estadística de la población estudiantil. Se determinará variables a medir en encuestas diseñadas con el objeto de producir conocimiento que pueda aplicarse para mejorar la calidad de vida y salud de la población estudiantil. Se aplicarán las técnicas del AMD para analizar variables del SIU Guarani, a fin de caracterizar a la población estudiantil de las facultades involucradas. También se analizará las

variables del SIU Kolla a fin de caracterizar a la población de graduados de las facultades involucradas e identificar, entre otras, sus necesidades de capacitación y/o actualización.

Para realizar la investigación que permitirá la elaboración y el análisis de las tipologías de los estudiantes universitarios, tanto en su fase de tránsito como egreso de la universidad, se llevará a cabo un análisis de datos, utilizando métodos exploratorios multidimensionales que se inscriben en la corriente francesa de análisis de datos (Benzécri, 1976; Lebart, Morineau y Oiron, 1995). Estas técnicas son especialmente adecuadas para el tratamiento de datos de encuestas y permiten el análisis de la estructura de la información, sin especificación de modelos estadísticos a priori, analizando todas las variables en forma simultánea en espacios a través del recurso de representaciones gráficas en planos factoriales (Moscoloni, 1999).

Siguiendo con esta línea de análisis exploratorio de datos se incorpora la metodología de creación de objetos simbólicos (E. Diday, 1997), que tiene como punto de partida consultas a una base de datos relacional. Por medio de estas consultas se agregan individuos con características comunes, como por ejemplo, carreras, residencia, situación laboral, etc. Los objetos simbólicos creados son también almacenados en tablas, llamadas tablas simbólicas. Cada celda de estas tablas, con objetos simbólicos por filas y variables por columnas, puede contener datos de diferentes tipos.

En este proyecto se pretende utilizar los objetos simbólicos para definir grupos de estudiantes con características comunes, es decir para caracterizar desde distintos enfoques, según el análisis, a estudiantes universitarios.

Por otra parte, para el análisis confirmatorio, se utilizará el modelado estocástico con la aplicación de modelos loglineales y modelos probabilísticos gráficos.

Hasta el momento se cuentan con los siguientes resultados que favorecerán la concreción de lo

anteriormente expuesto.

- *"Aplicación de técnicas multivariadas para la caracterización del ingresante a Medicina de la Universidad Nacional de Córdoba (Argentina)"* XXXIII Coloquio Argentino de Estadística. Universidad Nacional de Córdoba. Octubre 2005
- *"Explorando aspectos del conocimiento probabilístico en estudiantes preuniversitarios, con y sin formación previa."* Octavo Congreso Latinoamericano Argentino de Sociedades Estadística (IIIV CLATSE). Facultad de Ciencias Económicas y de Administración. Universidad de la República. Montevideo, Uruguay. Octubre 2008
- *"Análisis descriptivo uni y multivariado para caracterizar la variabilidad del razonamiento probabilístico en alumnos de nivel secundario."* XXXIV Reunión de Educación Matemática (REM. Universidad Nacional de Tucumán, Facultad de Ciencias Exactas, en la ciudad de San Miguel de Tucumán. Setiembre de 2011.
- *"Análisis Multivariado para identificar concepciones sobre evaluación"*, en las Terceras Jornadas de Ingreso y Permanencia en carreras Científico-tecnológicas". Universidad Nacional de San Juan, Facultad de Ingeniería. Mayo de 2012
- *"Modelos Log-lineales Gráficos y su aplicación en un estudio sobre razonamiento probabilístico."* X Congreso Latinoamericano de Sociedades de Estadística (X CLATSE). Universidad Nacional de Córdoba. Octubre de 2012.

- *"Análisis de asociaciones entre variables categorizadas en un*

*estudio sobre razonamiento probabilístico.* LXII Reunión Anual de Comunicaciones Científicas Rosario. Septiembre de 2013

- Grupo de discusión Red de colaboración en investigación de la educación estadística con la representación de *La Línea de investigación y grupo de investigación en Argentina.* Vigésimo Séptima Reunión Latinoamericana de Matemática Educativa (RELME 27). Buenos Aires, Julio de 2013.

Trabajo de tesis para acceder al grado de Magíster en Estadística aplicada desarrollado por la Mg. Ana María Ruiz, denominado *Representación gráfica y modelado de asociaciones múltiples: su aplicación en el análisis de heurísticas y sesgos del razonamiento probabilístico de alumnos de nivel secundario.* Abril 2011.

## Formación de Recursos Humanos

Tres integrantes del equipo trabajarán en sus tesis de posgrado. Uno en una tesis de maestría en Estadística, Otro en una maestría en Informática y otro en una de doctorado en Computación.

## Referencias

1. ALBATCH, P. (2001): Educación superior comparada. El conocimiento, la universidad y el desarrollo. Cátedra Unesco de Historia y Futuro de la Universidad. Colección Educación Superior. Universidad de Palermo. España.
2. ATO, M.; LÓPEZ, J.A. (1996): Análisis estadístico para datos categóricos. Madrid:

3. BENZÉCRI, Jean Paul (1976) : L'Analyse des données, T.I La taxonomie T.II L'Analyse des correspondances. Dunod. París.
4. CHIROLEU, Adriana (1999): El ingreso a la universidad. Las experiencias de Argentina y Brasil. UNR Editora.
5. DIDAY, Edwin (1992): Análisis de datos y clasificación automática numérica y simbólica. EUSTAT, Vitoria-Gasteitz. (1997): Análisis de datos simbólicos. Ed. IRICE, Rosario.
6. DIDAY, Edwin, y LECHEVALLIER, Yves Symbolic (1991): Numeric data analysis and learning, Versailles, September 18-20. INRIA, Nova Science Publishers Inc. New York.
7. FERNÁNDEZ AGUIRRE, Karmele: IV International Meeting of Multidimensional Data Analysis (NGUS'97), Bilbao, September 10-12, 1997. Universidad del País Vasco, Bilbao.
8. GARCÍA DE FANELLI, Ana María (2005): Acceso, abandono y graduación en la educación superior argentina. SITEAL, Debate 5. Disponible en internet: <http://www.siteal.iipe-oei.org/> [consulta: setiembre 2005].
9. LEBART, Ludovic; MORINEAU, Alain, y PIRON, Marie (1995): Statistique exploratoire multidimensionnelle. Dunod. París.
10. MOSCOLONI, Nora (2005): Las nubes de datos. Métodos para analizar la complejidad. UNR Editora, Rosario.