

USO DE ANÁLISIS MULTIVARIADO PARA CLASIFICAR A LOS ALUMNOS INGRESANTES EN EL ÁREA MATEMÁTICA SEGUN SU FORMACION PREVIA AL INGRESO A LA UNIVERSIDAD

Autores: Olga ÁVILA *^{,1}
Eleonora CERATI *^{,2}
Roberto MACÍAS *^{,§,3}
Claudia REDOLATTI *^{,4}
Ingrid SCHWER *^{,5}
María Laura TAVERNA *^{,6}

(*) Depto. de Matemática, Facultad de Ingeniería Química, UNL, Santa Fe,
(§), IMAL, CONICET, Santa Fe,
(1) oavila@imalpde.ceride.gov.ar
(2) ecerati@math.unl.edu.ar
(3) rmacias@math.unl.edu.ar
(4) clbere2001@yahoo.com
(5) ischwer@fiqus.unl.edu.ar
(6) ltaverna@fiqus.unl.edu.ar

“Área temática:” Investigación Educativa

“Palabras claves:” análisis discriminante, regresión logística, ingresante, rendimiento, matemática

INTRODUCCION

A partir del año 2000 se realiza un seguimiento estadístico para evaluar distintos aspectos del rendimiento en matemática de los alumnos ingresantes de las diferentes carreras de la Facultad de Ingeniería Química (FIQ) dependiente de la Universidad Nacional del Litoral (UNL), con sede en la ciudad de Santa Fe. Se dictan, entre otras, las carreras de Ingeniería Química (IQ), Licenciatura en Química (LQ), Ingeniería Industrial (II), Analista Industrial (AI), Ingeniería en Alimentos (IA), Químico Analista (QA) y Profesorado en Química (PQ). Estas carreras cuentan con un núcleo de materias básicas, del cual forma parte el primer curso de Matemática, llamado Matemática A. Esta materia abarca contenidos de cálculo y álgebra lineal.

A partir del año 2001 es un requisito obligatorio la aprobación del Curso de Articulación Disciplinar en Matemática para el cursado de Matemática A. Los ingresantes que no aprueban dicho curso deben asistir a los Cursos de Articulación Disciplinar tutoriales durante el primer cuatrimestre de la carrera. De esta forma los alumnos ingresantes quedan separados en dos grupos: aquellos que deben realizar el Curso Tutorial de Matemática ya que necesitan de una mayor formación previa en los contenidos del nivel Polimodal y los que tienen la formación suficiente para realizar Matemática A.

Estudios realizados en trabajos anteriores con alumnos ingresantes en las carreras de la Facultad mostraron que la aprobación del Curso de Articulación en el área matemática es un factor importante y decisivo para un buen desempeño de los alumnos en el primer curso de Matemática [1]. Por este motivo surge la necesidad de estudiar en mayor detalle distintos aspectos que diferencian a los alumnos que aprobaron este curso de aquellos que deben realizar el Curso Tutorial de Matemática para reforzar contenidos del polimodal.

El objetivo de este trabajo es describir resultados obtenidos de la utilización de las técnicas multivariadas de análisis discriminante y regresión logística para distinguir entre estos dos grupos. Ambos análisis permiten identificar un conjunto de variables independientes de un grupo mayor, como las variables mejores predictoras.

Datos

Para la obtención de los datos se utilizaron los formularios SUR, los resultados obtenidos del Curso de Articulación Disciplinar en Matemática, y la implementación de una encuesta ya realizada en el año 2000 donde se preguntan cuestiones relacionados con el curso de articulación disciplinar, y se realizan preguntas referidas al estudio de recta, parábola y

trigonometría que permiten reflejar aspectos de la formación obtenida en el nivel polimodal en el área matemática.

Variables

Inicialmente se consideran las siguientes variables:

- a) Tuvo Matemática en el último año del nivel polimodal (variable dicotómica)
- b) Estudió recta (variable dicotómica)
- c) Estudió parábola (variable dicotómica)
- d) Estudió trigonometría (variable dicotómica)
- e) Situación previa: si cuenta con estudios universitarios o terciarios previos (variable dicotómica)
- f) Cantidad de conocimientos nuevos que proporciona el Curso de Articulación: se han considerado tres modalidades ninguno, pocos y muchos
- g) Dependencia del establecimiento donde cursó el polimodal (variable dicotómica)
- h) Orientación del polimodal (variable dicotómica)
- i) Nota del examen del Curso de Articulación en Matemática (variable continua)
- j) Situación laboral: se han considerado las modalidades no trabaja, menos de 20 horas, entre 20 y 30 horas, y más de 30 horas.

Salvo la variable i), las restantes se obtienen de la encuesta mencionada.

Análisis Descriptivo de las variables que definen los modelos

El Software utilizado para el procesamiento de los datos fue el SPSS versión 7.0.

Estudió Parábola

	Frecuencia
no	27
si	212
Total	239

Estudió Trigonometría

	Frecuencia
no	40
si	199
Total	239

Matemática Último año

	Frecuencia
,00	88
si	151
Total	239

NOTA APOYO

	Frecuencia
1,00	2
2,00	10
3,00	7
4,00	34
5,00	13
6,00	24
7,00	19
8,00	19
9,00	22
10,00	48
Total	198

Análisis Discriminante

La técnica multivariada de análisis discriminante permite relacionar diferentes aspectos relevantes para discriminar en dos grupos a los fines de caracterizar la formación de los ingresantes en la Universidad. Es necesario tener grupos excluyentes y totalmente diferenciales. En este caso son dos: aquellos que están en condiciones de cursar Matemática A y quienes deben asistir al Curso de Apoyo para reforzar los conocimientos impartidos en el polimodal.

Para utilizar esta técnica se necesita la información relacionada con dos aspectos esenciales: un conjunto de casos a los que se les conozca el grupo de pertenencia y un conjunto de variables que identifiquen las características que mejor miden las diferencias entre esos grupos.

El análisis discriminante [2] permite construir una función que es una combinación lineal de variables que evaluada en cada caso proporciona un puntaje que permite asignarlo al grupo de pertenencia. Una vez que esta función discrimina correctamente los casos conocidos pueden utilizarse para clasificar nuevos casos de los que se desconoce el grupo de pertenencia.

Debido a una gran diferencia en cantidad de datos disponibles en los dos grupos estudiados, se utilizan probabilidades proporcionales al tamaño de los grupos para hacer el análisis.

Una etapa importante en la técnica de análisis discriminante para construir un modelo adecuado es poder seleccionar correctamente un subconjunto de las variables que resulten significativas.

Del conjunto de variables inicialmente consideradas, se seleccionan un subconjunto de ellas; para ellas sus medias (que se muestran en la tabla 1) resultan significativamente diferentes a un nivel de 5% :

- a) Matemática en el último año.
- b) Estudió parábola
- c) Estudió trigonometría
- d) Nota del examen del Curso de Ariculación en Matemática

Inicialmente se realiza un test de igualdad de medias entre grupos a todas las variables consideradas. En la tabla 2 se muestran sólo aquellas que indican diferencias estadísticamente significativas a un nivel del 5 %. Estas variables resultan ser las apropiadas para construir el modelo buscado.

	Lambda de Wilks	F	G1	G2	Significación
Estudió Parábola	0.948	11,022	1	196	0.001
Estudió Trigonometría	0.964	7.336	1	196	0.007
Nota Apoyo	0.676	93.95	1	196	0.000
Mat. Ultimo año	0.96	8.13	1	196	0.005

Tabla 1: Pruebas de igualdad de las medias de los grupos

Existen otros estadísticos propios de la técnica como el coeficiente de correlación canónica (0,593) y Lambda de Wilks (0,649) que validan el modelo elegido de todos los propuestos.

Función discriminante

La Tabla 2 detalla los coeficientes estandarizados de la función obtenida con las variables que finalmente resultaron significativas. El valor absoluto de estos coeficientes indican la importancia relativa de la variable correspondiente en el cálculo de la función discriminante.

	Coeficientes
Nota Apoyo	,904
Estudió Parábola	,294
Estudió Trigonometría	,116
Matemática Ultimo año	-,089

Tabla 2: Coeficientes estandarizados de las funciones discriminantes canónicas

La función queda definida de la siguiente manera:

$$D_i = 0,294 \text{ estudióParábola} + 0,116 \text{ Estudió Trigonometría} + 0,904 \text{ Nota Apoyo} - 0,089 \text{ Mat. Ultimo año}$$

Esta función evaluada en cada caso arroja un puntaje D_i , en base al cual se establece el grupo al que pertenece. El programa calcula estos puntajes para cada caso.

La Tabla 3 de resumen, muestra la cantidad de datos discriminados correcta e incorrectamente. En este estudio se tiene que el 96 % se clasifica correctamente.

		Grupo de pertenencia pronosticado		Total
discriminante		no	si	
Original	Recuento	no	si	
		14	5	19
		3	176	179
%		73,7	26,3	100,0
		1,7	98,3	100,0

Tabla 3: Resultados de la clasificación

Como se puede observar de la tabla, en el grupo correspondiente a alumnos del Curso Tutorial la técnica es menos eficiente en su discriminación (26,3 % de casos mal clasificados en el grupo de alumnos del Curso Tutorial). Una de las posibles causas es la diferencia marcada en los tamaños de los grupos. No obstante, se debe aclarar que en el grupo minoritario no se obtuvo respuesta en algunas de las variables, y como se dijo anteriormente la técnica requiere casos con información completa. Se desea repetir este estudio el siguiente año, poniendo especial esfuerzo en obtener datos en el grupo de alumnos que no aprueban el Curso de Articulación Disciplinar. Para clasificar nuevos casos se utilizan funciones de clasificación para cada grupo cuyos coeficientes se presentan en la tabla 4. Cuando se utilizan estas funciones, se clasifica un individuo en el grupo para el que la función sea mayor.

Variable	Coeficiente en la función discriminante	
	Alumno del Curso Apoyo	Alumno de Matemática A
Estudió Parábola	11.97	14.80
Estudió Trigonometría	3.63	4.43
Nota Apoyo	0.58	1.62
Mat. último año	5.49	5.02
Constante	-11.16	-16.51

Tabla 4: Coeficientes para la función discriminante lineal

Si bien son muy pocos los casos mal clasificados de los alumnos que pertenecen al grupo de Matemática A y el modelo los clasifica como pertenecientes al grupo Tutorial (2 %), la lectura de esta situación nos es de utilidad para detectar casos que cuentan aún con deficiencias en su formación previa en matemática comprometiendo su desempeño en este curso. El uso de estrategias pedagógicas alternativas permitiría subsanar esta situación.

Regresión logística

El análisis de regresión logística presenta la ventaja de no requerir que las variables se distribuyan según la ley normal, lo que muchos argumentan como la razón fundamental para que este enfoque resulte mucho más robusto que el enfoque discriminante. Debido a que la mayoría de las variables independientes son categóricas en este caso, se presenta un modelo utilizando esta técnica.

Este tipo de modelo, también llamado modelo de respuesta cualitativa, tiene utilidad para pronosticar que sucederá cuando existen dos posibilidades, en nuestro caso, superar la asignatura o no superarla. Esto se logra mediante un modelo que calcula probabilidades para cada caso permitiendo asignarlo a un grupo bien definido según si esta probabilidad supera o no un punto de corte previamente definido. (0.65 en nuestro caso).

El modelo resultante es el siguiente:

$$P(Z = 1) = \frac{1}{1 + \exp(-112.84 + 32.41\text{NOTAP} - 0.45\text{MATUIT} - 0.12\text{ESTUDIOT} + 0.69\text{ESTUDIOP})}$$

Las pruebas de bondad de ajuste de Hosmer y Lemeshow y la prueba que considera $-2\log$ de la verosimilitud indican un buen ajuste del modelo aquí presentado.

Tabla de clasificación^a

Observado		Pronosticado		
		discriminante		Porcentaje correcto
		no	si	
discriminante	no	19	0	100,0
	si	0	179	100,0
Porcentaje global				100,0

a. El valor de corte es ,650

Tabla 5: Resultados de la clasificación

Conclusiones

Periódicamente se realizan estadísticas usando las bases de datos de la asignatura Matemática A a los fines de diagnosticar y tomar decisiones en busca de una mejor organización y calidad de enseñanza. Persiguiendo este objetivo es que se ha elegido aplicar técnicas de análisis multivariado que permiten obtener una clasificación de los alumnos ingresantes de nuestra facultad según su calidad en la formación matemática con la que se inician.

Tanto las técnicas de análisis discriminante como regresión logística proporcionaron buenas clasificaciones para los casos estudiados (96 y 100 % de los casos correctamente clasificados). Cuando la función discriminante aplicada a casos conocidos clasifica incorrectamente un alumno de matemática A y lo coloca en el grupo que debe realizar el curso tutorial, indica que debe dársele especial atención debido a que puede evidenciar deficiencia en su formación matemática.

Estas técnicas permiten realizar un análisis individual haciendo un diagnóstico sobre la formación actual de cada alumno, evaluando si es apropiada para un buen desempeño en el primer curso de matemática.

En el futuro, el agregado de más preguntas tendientes a obtener mayor información, permitirá una clasificación más desagregada. Considerando los alumnos que cursan la primer materia,

Matemática A, es posible aplicar las técnicas para discriminarlos en niveles de formación y aptitud en dicha área.

La detección de los grupos con deficiencias permiten implementar políticas y estrategias de enseñanza para dar solución a dichas deficiencias, reduciendo a su vez situaciones de fracaso.

“Referencias bibliográficas:”

[1] Avila O. y otros, Rendimiento en Matemática de los alumnos ingresantes en la Facultad de Ingeniería Química, Universidad Nacional del Litoral (Actas de Congreso, UMA) Salta, Argentina 2005..

[2] Johnson R. and D. Wichern, Statistical Multivariate Statistical Analysis, Fourth Edition, Prentice Hall, 1998.

[3] LeBold W. y otros, The use of discriminant analysis for optimal placement (ASEE Annual Conference Proceedings, Purdue University, W. Lafayette), Estados Unidos, 1989.

[4] LeBold W. y otros, Understanding of Mathematics and Science: Efficient Models for Student Assessments (Vol.41, Nro 1 IEEE Transactions on Education, Purdue University, W. Lafayette), Estados Unidos, 1998