

USO DE ANÁLISIS MULTIVARIADO PARA CARACTERIZAR LA FORMACIÓN MATEMÁTICA DE LOS ALUMNOS INGRESANTES Y MEDIR SU DESEMPEÑO EN LA PRIMERA ASIGNATURA DEL ÁREA EN LA UNIVERSIDAD

Autores: Olga ÁVILA *
Eleonora CERATI *
Roberto MACÍAS *
Claudia REDOLATTI *
Ingrid SCHWER *
María Laura TAVERNA *

(*) Depto. de Matemática, Facultad de Ingeniería Química, UNL, Santa Fe
(**) IMAL, CONICET, Santa Fe,
(*), oavila@imalpde.ceride.gov.ar
(*), eleonoracerati@gmail.com
(*),(**) Roberto.a.macias@gmail.com

Sección A

Palabras claves: análisis discriminante, regresión logística, ingresante, matemática

Introducción

Uno de los problemas más serios que enfrentan hoy en día las universidades en el dictado de sus primeros cursos es la heterogeneidad del nivel de conocimientos alcanzado por los alumnos en los estudios previos y las carencias con las que cuentan, dificultando el desempeño en materias como matemática. Por esta razón es importante estudiar los aspectos que pueden influir en el rendimiento de los alumnos ingresantes evitando situaciones de fracaso y favoreciendo la retención de los mismos

En la Universidad Nacional del Litoral (UNL), con sede en la ciudad de Santa Fe, a partir del año 2001 es un requisito obligatorio la aprobación del Curso de Articulación Disciplinar en Matemática para el cursado de la primer materia en matemática llamada Matemática A. Los alumnos que no logran aprobarlo deben asistir durante el primer cuatrimestre de su primer año a un curso de apoyo (CAM) con el fin de lograr una mejor formación necesaria para realizar Matemática A.

Desde el año 2000 se realiza un seguimiento estadístico a los alumnos ingresantes de las carreras de la Facultad de Ingeniería Química (FIQ) dependiente de la (UNL) para evaluar distintos aspectos de su rendimiento en el área matemática.

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]. En virtud de ello surge la necesidad de analizar 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 como así también analizar estos aspectos considerando solamente aquellos alumnos ingresantes que aprobaron Matemática A..

El objetivo de este trabajo es mostrar resultados obtenidos de utilizar las técnicas multivariadas de análisis discriminante y regresión logística en alumnos ingresantes. Ambos análisis multivariados permiten identificar un conjunto de variables independientes de un grupo mayor, como las variables mejores predictoras para clasificar los alumnos ingresantes en dos grupos: alumnos en condiciones de cursar Matemática A; y alumnos que deben realizar un curso tutorial como curso de apoyo para lograr una mejor formación en matemática previo al cursado de este primer curso. Un análisis discriminante adicional se realiza a los alumnos de la asignatura Matemática A con el objeto de detectar variables que diferencien aquellos alumnos que aprobaron la materia de los que no lo hicieron.

Datos

Para la obtención de los datos de los ingresantes 2007 se utilizaron los resultados obtenidos del Curso de Articulación Disciplinar en Matemática, y una encuesta similar a la realizada en el año 2000 donde se recogen datos relacionados con el curso de articulación disciplinar y se realizan preguntas referidas al estudio de recta, parábola y trigonometría. Las respuestas a estas preguntas permiten reflejar en líneas generales aspectos de la formación obtenida en el nivel polimodal en el área matemática.

Variables

Inicialmente se consideran las siguientes variables:

- Cursó Matemática en el último año del nivel polimodal (variable dicotómica).
- Recuerda haber estudiado recta (variable dicotómica).
- Recuerda haber estudiado parábola (variable dicotómica).
- Recuerda haber estudiado trigonometría (variable dicotómica).
- Situación previa: cuenta con estudios universitarios o terciarios previos
- Cantidad de conocimientos nuevos que proporciona el Curso de Articulación: se han considerado dos modalidades pocos y muchos.
- Dependencia del establecimiento donde cursó el polimodal con las siguientes modalidades: provincial, dependiente de la UNL, privada religiosa, privada particular, Instituto militar y otros.
- Orientación del polimodal con las siguientes modalidades: humanidades y ciencias sociales; economía y gestión de las organizaciones; ciencias naturales; producción de bienes y servicios; comunicación, arte y diseño y otros.
- Nota del examen del Curso de Articulación en Matemática (CAM)(variable continua).
- Situación laboral: se han considerado las modalidades no trabaja, 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.

Se presenta a continuación la tabla 1 de frecuencias asociadas a las variables que resultaron posteriormente significativas en los modelos estadísticos considerados en el presente trabajo.

APROBACION DEL CAM		Media	Desv. típ.	No ponderados
0	RECTA	,9130	,28810	23
	PARABOLA	,8261	,38755	23
	NOTA DEL CAM	3,3035	1,34645	23
	CERT.EDUC.POLIM.	2,0435	1,02151	23
	CONOC.ESC.MEDIA	1,6522	,48698	23
	TRIGONOMETRIA	,8261	,38755	23
	MAT ULTIMO AÑO	1,2609	,44898	23
	DEPENDENCIA	1,8261	1,07247	23
	SITUACION ESTUDIOS PREVIOS	,4348	,99206	23
	TRABAJA	1,5217	,94722	23
	1	RECTA	1,0000	,00000
PARABOLA		1,0000	,00000	84
NOTA DEL CAM		7,5158	1,21238	84
CERT.EDUC.POLIM.		3,5119	1,65372	84
CONOC.ESC.MEDIA		1,8571	,35203	84
TRIGONOMETRIA		,9881	,10911	84
MAT ULTIMO AÑO		1,3571	,48204	84
DEPENDENCIA		2,0000	1,09764	84
SITUACION ESTUDIOS PREVIOS		,1429	,51817	84
TRABAJA		1,1905	,56985	84
Total		RECTA	,9813	,13607
	PARABOLA	,9626	,19059	107
	NOTA DEL CAM	6,6104	2,13300	107
	CERT.EDUC.POLIM.	3,1963	1,65084	107
	CONOC.ESC.MEDIA	1,8131	,39168	107
	TRIGONOMETRIA	,9533	,21205	107
	MAT ULTIMO AÑO	1,3364	,47472	107
	DEPENDENCIA	1,9626	1,08962	107
	SITUACION ESTUDIOS PREVIOS	,2056	,65500	107
	TRABAJA	1,2617	,67763	107

Tabla 1: Estadísticos de las variables según grupos.

Análisis Discriminante

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

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

El análisis discriminante [3,4,5] 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 puede utilizarse para clasificar nuevos casos cuyo grupo de pertenencia se desconoce.

La técnica permite trabajar con grupos con una gran diferencia en cantidad de datos, utilizando probabilidades proporcionales al tamaño de los grupos, situación a tenerse en cuenta en el presente trabajo ya que se tienen solamente 23 alumnos que asistieron al Curso de Apoyo y 84 alumnos que estuvieron en condiciones de cursar Matemática A.

Una etapa importante en la técnica de análisis discriminante para construir un modelo adecuado es poder seleccionar correctamente un subconjunto de variables capaces de discriminar entre los dos grupos identificados. Para tal fin inicialmente se realiza un test de igualdad de medias entre grupos a todas las variables disponibles (Tabla 2).

	Lambda de Wilks	F	gl1	gl2	Sig.
RECTA	,930	7,850	1	105	,006
PARABOLA	,858	17,354	1	105	,000
NOTA DEL CAM	,336	207,808	1	105	,000
CERT. EDUC. POLIM.	,865	16,356	1	105	,000
CONOC. ESC. MEDIA	,953	5,138	1	105	,025
TRIGONOMETRIA	,901	11,593	1	105	,001
MAT ULTIMO AÑO	,993	,741	1	105	,391
DEPENDENCIA	,996	,458	1	105	,500
SITUACION ESTUDIOS PREVIOS	,966	3,677	1	105	,058
TRABAJA	,959	4,456	1	105	,037

Tabla 2: Prueba de igualdad de medias

Se observa que solamente las variables " Matemática último año ", " Dependencia " y " Situación estudios previos " no son significativas al nivel del 5 %.

Las variables restantes se utilizan para el estudio de un modelo discriminante. La técnica implementada indica que las variables consideradas previamente en los ítems b), c), h) e i) resultan las apropiadas para la construcción del modelo.

Existen otros estadísticos propios de la técnica, como el coeficiente de correlación canónica y Lambda de Wilks, que permiten elegir el modelo más apropiado de todos los modelos que pueden construirse a partir de las cuatro variables previamente seleccionadas. Ambos coeficientes miden las diferencias entre los grupos debidas a las funciones discriminantes. El estadístico Lambda de Wilks (Tabla 3) expresa la proporción de variabilidad total no debida a las diferencias entre los grupos (variabilidad dentro de los grupos); permite además contrastar la hipótesis nula de que las medias multivariantes de los grupos (centroides) son iguales. Al realizar esta prueba para nuestro caso se rechaza la hipótesis nula a una significación mayor al 1 %, lo cual indica que el modelo propuesto es apropiado para discriminar entre los grupos.

Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1	,288	195,173	4	,000

Tabla 3: Significación del estadístico Lambda de Wilks

El coeficiente de correlación canónica (Tabla 4) mide para la función discriminante, el grado en que difieren las medias de dicha función en los distintos grupos. Un valor alto del coeficiente indica una fuerte relación entre el grupo de pertenencia y los valores de la función discriminante. En este caso, resulta un valor moderado para este coeficiente.

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	2,466	100,0	100,0	,844

Tabla 4: Correlación canónica

Función discriminante

La Tabla 5 detalla los coeficientes estandarizados de la función obtenida con las variables que finalmente resultaron significativas. El valor absoluto de estos coeficientes indica la importancia relativa de la variable correspondiente en el cálculo de la función discriminante.

	Función
	1
RECTA	,337
PARABOLA	,442
NOTA DEL CAM	,937
ORIENTACION	
POLIMODAL	,196

Tabla 5: Coeficientes estandarizados de las funciones discriminantes canónicas

La función queda definida de la siguiente manera:

$$D_i = 0,337 \text{ estudió recta} + 0,442 \text{ estudió parábola} + 0,937 \text{ nota del CAM} + 0,196 \text{ orientación polimodal}$$

Esta función evaluada en cada caso arroja un puntaje D_i , que permite establecer el grupo de pertenencia del alumno.

			Grupo de pertenencia pronosticado		Total
			0	1	
Original	Recuento	0	27	0	27
		1	0	134	134
	%	0	100,0	,0	100,0
		1	,0	100,0	100,0

Tabla 6: Resultados de la clasificación

Para poder interpretar la Tabla 6 se debe recordar que se desea discriminar el grupo formado por alumnos que deben realizar el Curso Tutorial de Matemática es decir que no aprobaron el Curso de Articulación Disciplinar, del grupo de estudiantes que tienen la formación suficiente para realizar Matemática A o sea que han aprobado el Curso de Articulación Disciplinar. Esta Tabla de resumen, muestra la cantidad de datos discriminados correcta e incorrectamente. En este estudio se tiene que el 100 % se clasifica correctamente.

Fisher propuso una función de clasificación para cada grupo. En el caso de dos grupos la diferencia entre ambas funciones da lugar a un vector de coeficientes proporcional a los coeficientes no estandarizados de la función discriminante canónica. Para clasificar nuevos casos se utilizan estas funciones de clasificación para cada grupo cuyos coeficientes se presentan en la tabla 7. Al utilizarlas, se clasifica un alumno en el grupo para el que la función sea mayor.

	APROBACION DEL CAM	
	NO	SI
RECTA	93,857	106,896
PARABOLA	44,331	55,868
NOTA DEL CAM	4,159	7,435
ORIENTACION		
POLIMODAL	,748	1,266
(Constante)	-71,017	-111,628

Tabla 7: Coeficientes para la función discriminante lineal

Se realizó un análisis discriminante sin tener en cuenta la nota del CAM, resultando un modelo conformado por las variables "Orientación del Polimodal", las referidas a si recuerda haber estudiado en la escuela media temas como recta, parábola y trigonometría, indicando que este

tipo de variables son apropiadas para la discriminación de estos grupos.

Regresión logística

Para cada alumno se desea establecer si tiene una formación lo suficientemente robusta como para realizar el primer curso de las carreras de Ingeniería, Matemática A, o requiere del apoyo especial del Curso Tutorial en Matemática, quedando identificados los alumnos ingresantes en dos grupos. Estos dos grupos pueden verse como una variable dicotómica que considera la pertenencia o no de los alumnos a Matemática A según si aprobaron el Curso de Articulación Disciplinar. Un modelo apropiado cuando la variable dependiente es dicotómica es el modelo de regresión logística. Este tipo de modelo, también llamado modelo de respuesta cualitativa, tiene utilidad para pronosticar qué sucederá cuando existan dos posibilidades, en nuestro caso, superar dicho curso o no. 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 fijado (0.50 en nuestro caso).

El análisis de regresión logística [2,6] 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 notoriamente más robusto que el enfoque discriminante.

Este tipo de modelos permiten estimar o predecir la probabilidad de que un individuo posea una característica en función de determinadas cualidades individuales o variables (x_1, x_2, \dots, x_k).

En regresión logística el modelo matemático que mejor estima tal probabilidad, debido a que restringe los valores a su rango $0 < P < 1$ y aproxima a una forma "S" de la curva, es el siguiente:

$$P(Y = 1) = \frac{e^{a+bx}}{1 + e^{a+bx}} \text{ o equivalentemente: } P(Y = 1) = \frac{1}{1 + e^{-(a+bx)}}$$

Para la estimación de los coeficientes b utilizamos el método de máxima verosimilitud.

Para nuestros datos el modelo resultante es el siguiente:

$$P(Z = 1) = \frac{1}{1 + \exp(-129.962 + 21.919\text{NotCAM} + 16.932\text{EstRecta})}$$

Para ver la de bondad de ajuste del modelo se consideran las pruebas de Hosmer y Lemeshow y la prueba que considera $-2\log$ de la verosimilitud. Ambos test indican un buen ajuste del modelo aquí presentado. Al utilizar el modelo resultante se obtiene una clasificación de los casos como se muestra en la tabla 8.

		Pronosticado	
		APROBADO	
Observado	no	no	si
	APROBADO	no	23
	si	0	84

Tabla 8: Resultados de la clasificación

La técnica de regresión logística requirió de una menor cantidad de variables para su modelo, destacando que éstas son un subconjunto de las utilizadas en el análisis discriminante.

Uso de Análisis discriminante con los alumnos de la asignatura Matemática A

Se realiza un estudio adicional de análisis discriminante a los alumnos que cursan la materia Matemática A. El objetivo que se persigue es identificar las variables que diferencian a los alumnos que han aprobado de los que no lo logran en los turnos julio y agosto correspondientes a los únicos turnos comprendidos entre primer y segundo cuatrimestre de cursado.

Las 3 variables, Recuerda haber estudiado recta, parábola y trigonometría consideradas en el análisis discriminante anterior no se tienen en cuenta aquí debido a que tienen un único valor para todos los alumnos de nuestra nueva población, Matemática A.

De las variables restantes un test de igualdad de medias permite seleccionar aquellas que se pueden tener en cuenta como posibles variables discriminantes. Resultaron apropiadas con un nivel de significación de 0.05: Cantidad de conocimientos nuevos que proporciona el Curso de

Articulación, Nota del examen del Curso de Articulación en Matemática (CAM) y Situación laboral. Los coeficientes de correlación canónica y Lambda de Wilks indican que modelo entre todos los posibles es el más apropiado:

$$D_i = -0.165 \text{ trabaja} + 0,985 \text{ nota del CAM}$$

Para este modelo el coeficiente de correlación canónica es de 0.827 y el coeficiente de Lambda de Wilks es de 0.317 indicando un modelo válido para discriminar entre los dos grupos.

Al aplicar el modelo a los alumnos de Matemática A surge la siguiente tabla de clasificación:

			Grupo de pronosticad		Total
			0	1	
Origina	Recuent	0	42	1	43
		1	0	173	173
	%	0	97,7	2,3	100,0
		1	,0	100,0	100,0

a. Clasificados correctamente el 99,5% de los casos originales

Tabla 9: Resultados de la clasificación

Este modelo está indicando cómo influye la condición de trabajar o no en el rendimiento de la asignatura, en tanto el coeficiente que acompaña en el modelo a la variable nota de apoyo del CAM al ser positivo y alto con respecto al de la variable asociada a trabajo indica que aquellos alumnos con notas más altas en el curso de articulación disciplinar tienen una formación más sólida en el área matemática que los prepara para realizar en mejores condiciones la asignatura.

Conclusiones

Las técnicas de análisis multivariado son herramientas de gran utilidad e interpretadas junto con las estadísticas que se vienen realizando con bases de datos de los alumnos ingresantes resultan apropiadas para diagnosticar y tomar decisiones en busca de una mejor organización y calidad de enseñanza. Las técnicas de análisis multivariado logran identificar un conjunto de variables que caracterizan la formación de los alumnos ingresantes en el área matemática.

- Tanto las técnicas de análisis discriminante como de regresión logística proporcionaron buenas clasificaciones para los alumnos ingresantes ubicándolos en el curso tutorial o en Matemática A. Ambos análisis estarían indicando que la aprobación del curso de articulación en matemática puede ser considerada una herramienta apropiada de medición de la formación de los ingresantes en el área matemática. Las variables resultantes en el modelo de análisis discriminante evidencian diferencias en los grupos de ingresantes respecto a la orientación de la educación polimodal, no así respecto al tipo de escuela a la cual asistieron (pública o privada).
- Para realizar el análisis multivariado a los alumnos de Matemática A se consideraron las mismas variables de medición que en los estudios anteriores. Salvo la variable si trabaja o no, las restantes caracterizan el nivel de formación previo al ingreso universitario y no tienen en cuenta los conocimientos adquiridos en esta asignatura. El modelo resultante muestra que los contenidos evaluados en el curso de articulación disciplinar dan una formación esencial sólida necesaria para un buen desempeño del alumno en la asignatura Matemática A. La otra variable presente en el modelo, si trabaja o no, influye también en el rendimiento del alumno.

“Referencias bibliográficas”:

- [1] Ávila 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] Hosmer D. y Lemeshow, *Applied Logistic Regresión*, Estados Unidos, 1989.
- [3] Johnson R. y D. Wichern, *Statistical Multivariate Statistical Analysis*, Fourth Edition,

Prentice Hall, 1998.

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

[5] 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.

[6] Pampel F., *Logistic Regresión*, Estados Unidos, 2000.