



Innovation in research and engineering education:
key factors for global competitiveness

Innovación en investigación y educación en ingeniería:
factores claves para la competitividad global

PROPUESTA PARA LA PREDICCIÓN DEL RENDIMIENTO ACADÉMICO DE LOS ESTUDIANTES DE LA UNIVERSIDAD DEL ATLÁNTICO, BASADO EN LA APLICACIÓN DEL ANÁLISIS DISCRIMINANTE

Adel Alfonso Mendoza Mendoza, Roberto José Herrera Acosta

Universidad del Atlántico
Barranquilla, Colombia

Resumen

En este artículo se propone que la Universidad del Atlántico implemente el uso del análisis discriminante, que es un modelo estadístico multivariado que tiene como objetivo encontrar la combinación lineal de las variables independientes que mejor permite diferenciar (discriminar) a los grupos. Una vez encontrada esa combinación (la función discriminante) podrá ser utilizada para clasificar nuevos casos. En este caso se pretende determinar qué factores discriminan (diferencian) el grupo de estudiantes que culminan con éxito un periodo académico (o una asignatura), del grupo de estudiantes que lamentablemente fracasan en el alcance de logros y finalmente no aprueban el periodo académico. La validación de este modelo dependerá del grado de precisión al escoger variables obligatoriamente de tipo cuantitativo que maximicen la función discriminante y pueda proporcionar información útil y confiable.

Palabras clave: análisis discriminante; rendimiento académico

Abstract

This paper proposes that the Universidad of Atlántico implement the use of discriminant analysis, a multivariate statistical model that aims to find the linear combination of independent variables that best allows you to differentiate (discriminate) groups. Once you have found this combination (discriminant function) may be used to classify new cases. In this case seeks to determine what factors discriminate (differentiate) the group of students who successfully complete a semester (or assignment), of the group of students who unfortunately fail in the scope of academic achievement. Validation of this model will depend on the degree of accuracy in choosing necessary quantitative variables that maximize the discriminant function and can provide useful and reliable.

Keywords: *discriminant analysis; academic performance*

1. Introducción

El rendimiento estudiantil es un concepto complejo, que aborda distintos aspectos, tanto individuales como colectivos, lo cual condiciona el aprendizaje y la acción académica. En este sentido aborda la actuación de los estudiantes, del docente, así como de los órganos directivos, de la infraestructura y cultura organizacional, por lo cual el rendimiento estudiantil no solo es una calificación y viene concatenada en una política de admisión (Bonucci, 1997), lo que conlleva a afirmar que son muchas las variables inherentes en un proceso de aprendizaje que en la medida que sean identificadas con un aceptable índice de precisión pueden ser manejadas y controladas para mejorarlas. El rendimiento académico no es el producto de una sola capacidad sino que es el resultado de varios factores que interactúan en la persona que aprende (Gómez et al, 2011)

En la actualidad la medición del rendimiento académico y su predicción a partir de los factores que lo puedan determinar está siendo estudiada por muchos investigadores y se ha convertido en una prioridad en las instituciones de educación superior (Guzmán 2012).

2. Justificación

Los diferentes programas académicos en la Universidad Atlántico, al igual que otras Instituciones de Educación Superior, registran desde hace algunos años, un elevado índice de deserción y una de las principales causas es el bajo rendimiento académico especialmente en las asignaturas de los primeros semestres por lo cual es importante determinar las principales características de este tipo de estudiantes para cada uno de los diferentes programas. Por lo anterior se propone la aplicación de un modelo discriminante que permita identificar las variables que incidan en el rendimiento académico (aprobar o desaprobado) y que posibilidad tiene un estudiante de pertenecer a uno de los dos grupos. El modelo es una herramienta predictiva, una vez que esta función discrimina correctamente los casos conocidos puede utilizarse para clasificar nuevos casos cuyo grupo de pertenencia se desconoce y diagnosticar la formación actual de cada alumno permitiendo a las directivas de la Universidad tomar acciones preventivas para aumentar el número de estudiantes que logren los objetivos académicos en cada uno de los programas.

El estudiante que ingresa a cualquier programa planea terminar su carrera en un tiempo normal de cinco años, pero no todos logran este objetivo, debido a factores que afectan su buen rendimiento académico que conlleva a repetir las asignaturas, que les adiciona tiempo y esfuerzo, prolongando su estadía académica en la universidad. A su vez, la Universidad propende por el mejoramiento continuo de los programas pero la realidad actual de fallas académicas es un obstáculo en la consecución de objetivos de calidad educativa. ¿Qué medidas tomar?

Considerando que la mayor deserción entre los estudiantes en las instituciones de educación superior en Colombia ocurre en los primeros semestres según el sistema para la prevención de la deserción de la educación superior del Ministerio de Educación Nacional (SPADIES 2012) con el presente proyecto se pretende crear una herramienta basada en el Análisis Discriminante Multivariado que permita predecir el rendimiento académico de los estudiantes, con el fin de tomar acciones oportunas de seguimiento y acompañamiento para así disminuir la cantidad de estudiantes que desertan en la Universidad del Atlántico.

3. Referente teórico

El concepto de análisis discriminante ha sido utilizado ampliamente en el estudio del desempeño estudiantil en muchas instituciones a nivel mundial como se puede visualizar en la problemática estudiada por Jiménez Fernández (1987). En el cual plantea el rendimiento escolar en un entorno universitario peculiar como es la universidad a distancia (UNED), caracterizado por un bajo rendimiento escolar y un elevado índice de abandono. Los datos con los que realiza su investigación «ex-post-facto» son de tipo académico (curso, número de asignaturas matriculadas, carrera, calificación obtenida, etc.) y de tipo socio personal (edad, sexo, desempeño o no de actividad profesional, tipo de actividad desempeñada en su caso, etc.).

Otros estudios se han centrado en la predicción del rendimiento académico con el objetivo de mejorar la selectividad o de encontrar una mejor alternativa para seleccionar a los alumnos que con mayor éxito puedan realizar estudios universitarios (por ej. Cuxart, Martí, Ferrer, 1997; Goberna, López y Pastor, 1987). Goberna *et al* (1987) buscan un criterio para la selección de los alumnos que deben acceder a la universidad, partiendo de lo inadecuado de utilizar la nota de selectividad o pruebas de aptitud de acceso a la universidad (PAAU) como criterio de acceso, puesto que la nota PAAU sin promediar con el expediente de bachillerato sólo explicaría el 6% del rendimiento posterior. El problema de encontrar buenos predictores del rendimiento futuro de manera que se reduzca el fracaso académico en los programas de postgrado ha recibido una especial atención en EEUU como nos indican Wilson y Hardgrave (1995), en un estudio en el que intentan encontrar la mejor técnica para la predicción del rendimiento académico, con el fin de aumentar la tasa de éxito en un programa de postgrado MBA (Masters Business Administration»). En la investigación de Wilson y Hardgrave (1995) se utilizaron como predictores del rendimiento académico distintos factores como: calificación media durante la carrera, calificación examen de admisión al postgrado, las cartas de recomendación, Experiencia profesional, entre otras. García *et al* (2000) buscan evaluar la capacidad de la regresión lineal y de la regresión logística en la predicción del rendimiento y del éxito/fracaso académico, partiendo de variables, como la asistencia y la participación en clase. Bacallao *et al* (2004) realizaron la construcción de un algoritmo para detectar estudiantes con alto riesgo de fracaso académico e identificar los mejores predictores del rendimiento dentro de los cuales se destacan el índice académico del preuniversitario, una prueba de ortografía, pruebas de nivel de entrada, exámenes de ingreso y variantes de diversas pruebas psicométricas. Moral (2006) estudia la relación del rendimiento académico con capacidad intelectual, alexitimia como indicador de inteligencia emocional, salud mental y variables socio demográficas. Reyes *et al* (2007) presentan una metodología para realizar predicciones del éxito en el primer semestre de los estudiantes de la carrera de Ingeniería Aplicando un modelo de regresión logística comparado los resultados con los del método de análisis discriminante.

Como parte de esta propuesta se pretende conocer como cada programa de la Universidad Atlántico pueda determinar qué variables son las mejores predictoras del éxito/fracaso de los estudiantes en el primer semestre (o asignaturas más importantes del primer semestre), teniendo en cuenta los estudios llevados previamente.

4. Conceptos básicos de análisis discriminante

El Análisis Discriminante es una técnica estadística multivariante, fue introducido por R. A. Fisher (1936) en el primer tratamiento moderno de problemas separatorios, cuya finalidad es analizar si existen diferencias significativas entre grupos de objetos respecto a un conjunto de variables medidas sobre los mismos para, en el caso de que existan, explicar en qué sentido se dan y proporcionar procedimientos de clasificación

sistemática de nuevas observaciones de origen desconocido en uno de los grupos analizados (Dallas, 2000). La variable dependiente de clasificación es una variable No Métrica, mientras que las variables independientes se supone que son métricas (Pérez, 2009)

Los objetivos primarios del análisis discriminante son la descripción de las diferencias entre grupos y la predicción de pertinencia a los grupos (Mileris, 2010). La interpretación de las diferencias entre los grupos responde al objetivo de determinar en qué medida un conjunto de características observadas en los individuos permite extraer dimensiones que diferencian a los grupos y cuáles de estas características presentan mayor poder de discriminación. El otro objetivo consiste en determinar una o más ecuaciones matemáticas, funciones discriminantes, que permitan la clasificación de nuevos casos a partir de la información que se tiene de ellos.

Las variables utilizadas en este análisis estadístico son denominadas variables discriminantes. Estas deben ser medidas en la escala de intervalo o razón para que las medidas y varianzas puedan ser calculadas e interpretadas. Un requerimiento para la utilización del Análisis Discriminante es que el número de casos observados (n) debe exceder por más de dos, al número de variables. Ninguna variable original puede ser combinación lineal de otras variables discriminantes ya que se tendría una redundancia en la información. Una combinación lineal es la suma de una o más variables que pueden haber sido ponderadas por términos constantes. Del mismo modo, dos variables que están perfectamente correlacionadas no pueden ser usadas al mismo tiempo.

Para la construcción del modelo se puede utilizar diferentes metodologías. Para la presente propuesta se recomienda utilizar la metodología de estimación por etapas para de esta manera ir seleccionando secuencialmente la variable que mejor discrimina en cada paso, de esta forma se identifica un conjunto reducido de variables y se eliminan las variables que no son útiles para discriminar entre los grupos. Como criterio de selección de variables que mejor discriminan se propone utilizar la distancia D^2 de Mahalanobis que es una medida de distancia generalizada y se basa en la distancia euclídea al cuadrado generalizada que se adecúa a varianzas desiguales, la regla de selección en este procedimiento es maximizar la distancia D^2 de Mahalanobis. La distancia multivariante entre los grupos a y b se define como:

$$H_{ab}^2 = (n - g) \sum_{i=1}^p \sum_{j=1}^p W_{ij} * (X_i^{(a)} - X_i^{(b)}) (X_j^{(a)} - X_j^{(b)})$$

donde n es el número de casos válidos, g es el número de grupos, $X_i^{(a)}$ es la media del grupo a en la i-ésima variable independiente, $X_i^{(b)}$ es la media del grupo b en la i-ésima variable independiente, y W_{ij} es un elemento de la inversa de la matriz de varianzas-covarianzas intra-grupos.

Como cualquier otra técnica estadística la aplicación de la misma ha de ir precedida de una comprobación de los supuestos asumidos por el modelo. El análisis discriminante se apoya en los siguientes supuestos:

- *Normalidad Multivariante.*
- *Igualdad de Matrices de Varianza-Covarianza.*
- *Linealidad.*
- *Ausencia de multicolinealidad y singularidad*

5. Metodología

Se propone aplicar la técnica de Análisis Discriminante como técnica de predicción de que un estudiante tenga éxito (Aprobar o Reprobar) el primer semestre de cualquier programa de la Universidad Atlántico.

Para la implementación de esta metodología en cada uno de los programas de la Universidad Atlántico se procederá de la siguiente manera:

Fase 1: Diseño de la metodología del análisis discriminante

Actividades:

- Caracterización de los estudiantes del programa para cuantificar el éxito académico (Aprobar o Reprobar) el primer semestre
- Diseño de encuestas para determinar que variables influyen en el rendimiento académico de los estudiantes en el programa académico respectivo, matriculados desde segundo hasta último semestre.
- Acceder a los sistemas de información nacional ICFES y al sistema de admisión y registros de la Universidad para analizar y estimar las variables identificadas en la recolección de datos.

Fase 2: Diseño del modelo análisis discriminante

Actividades:

- Clasificar y calificar las variables identificadas en el proceso de recolección de datos muestrales, por medio de una escala unificada.
- Aplicar herramientas informáticas y estadísticas para plantear el modelo discriminante.

Fase 3: Validación del modelo discriminante en nuevos casos

Actividades:

- Validar el resultado arrojado por el modelo con el resultado obtenido de éxito (aprobar o Reprobar) de los estudiantes de segundo a último semestre
- Utilizar el modelo discriminante con los valores de las variables aplicadas a cada estudiante de primer semestre para conocer con anticipación, bajo el supuesto de quien aprueba o desaprueba el primer semestre

6. Caso de aplicación

Se realizó una prueba preliminar simplificada de aplicación de Análisis Discriminante para predicción del rendimiento académico de los estudiantes de la asignatura de “*Química General y Laboratorio*” de un programa de Ingeniería durante tres periodos académicos.

Las variables de discriminación se tomaron del sistema de información nacional del ICFES:

- Puntaje de Química
- Promedio ICFES de todas las áreas
- Promedio ICFES de ciencias básicas
- Clasificación del plantel

Es importante asegurar la aleatoriedad en la selección de la muestra para que cualquier ordenación de las observaciones no afecte los procesos de estimación del modelo. La muestra tomada para la prueba fue de 54 estudiantes con un nivel de confianza del 95% y una precisión del 5%, esta se tomó de forma aleatoria

de una población de 162 estudiantes que cursaron la asignatura en los periodos académicos antes mencionado.

Los estudiantes se clasificaron en dos grupos de acuerdo a la nota definitiva en “Química General y Laboratorio” así:

Los del grupo 1 son aquellos cuya nota fue igual o superior a 3.5

Los del grupo 2 son aquellos cuya nota fue igual o inferior a 3.4¹

A los colegios de los cuales provienen los estudiantes se le asignó una calificación de 1 a 7 de acuerdo a la clasificación del mismo en las pruebas ICFES así:

Clasificación del Colegio en el ICFES	Puntaje Asignado
Muy Inferior	1
Inferior	2
Bajo	3
Medio	4
Alto	5
Superior	6
Muy Superior	7

Los resultados obtenidos para los estudiantes con los que se va a realizar el estudio preliminar se muestran en la siguiente tabla.

Estudiante	DATOS PRUEBAS ICFES				Nota Qca General. Universidad Atlántico	GRUPO
	Química	Promedio ICFES	Promedio de ciencias básicas	Clasificación Colegio		
1	52.16	51.16	49.87	6	3.8	1
2	50.94	51.52	52.70	7	3.8	1
3	52.59	48.89	47.87	5	4.3	1
4	52.59	52.50	48.64	5	4.3	1
5	50.50	51.93	50.07	5	3.6	1
6	53.84	57.09	53.29	6	4.1	1
7	57.36	62.82	59.47	7	4.7	1
8	59.25	52.76	50.66	7	3.5	1
9	52.59	53.84	52.14	6	4.1	1
10	56.05	47.07	46.42	5	4.2	1
11	49.30	48.60	48.30	5	4.1	1
12	39.51	38.52	39.93	4	3.8	1
13	43.56	47.35	50.01	5	3.5	1
14	46.00	46.44	47.14	5	3.6	1
15	47.22	47.02	45.37	3	4.2	1
16	46.00	45.33	45.91	7	4.2	1
17	42.03	45.25	44.81	6	3.7	1
18	47.13	45.05	42.54	4	3.4	1

¹ De acuerdo al seguimiento durante el semestre estudiantes con notas inferiores a 3.4 en esta asignatura presentan un alto grado de dificultad para aprobar el semestre.

19	50.94	50.34	53.91	4	4.4	1
20	55.57	54.06	54.20	6	4.2	1
21	52.16	52.18	48.19	6	4.3	1
22	56.05	54.19	57.82	6	4.8	1
23	45.38	45.88	41.60	3	4.5	1
24	52.59	46.20	46.89	7	3.5	1
25	47.13	44.82	47.76	4	3.9	1
26	54.59	54.73	53.43	5	4.6	1
27	56.08	49.55	48.99	6	3.6	1
28	56.40	52.03	53.66	7	3.7	1
29	43.72	44.83	44.94	5	3.5	1
30	54.30	52.49	52.80	6	3.5	1
31	60.27	52.57	52.93	5	4.5	1
32	54.59	51.25	55.88	7	3.5	1
33	62.42	59.06	60.80	3	4.3	1
34	55.27	45.34	50.59	6	3.8	1
35	42.49	48.21	47.19	4	3.0	2
36	52.59	49.54	51.07	6	3.0	2
37	43.56	42.44	43.53	4	3.3	2
38	48.83	44.94	45.43	4	1.3	2
39	47.67	48.59	47.99	4	3.4	2
40	52.59	45.79	48.93	6	3.4	2
41	48.83	49.06	49.45	6	3.4	2
42	42.49	38.93	37.47	3	3.0	2
43	57.31	61.98	59.85	7	3.0	2
44	50.50	46.13	47.72	4	3.0	2
45	39.03	44.11	41.60	2	3.0	2
46	41.62	38.48	36.79	3	3.0	2
47	36.34	37.47	42.00	5	3.1	2
48	40.58	42.39	38.85	5	3.4	2
49	43.56	38.06	41.93	5	3.0	2
50	47.62	49.62	45.51	5	3.4	2
51	43.72	42.99	40.32	4	2.2	2
52	54.30	44.23	46.03	5	3.4	2
53	47.38	51.71	52.95	6	3.1	2
54	47.38	43.55	43.88	4	3.3	2

Tabla 1. Resultados Pruebas ICFES y Notas de Química General y Lab.

Para la validación de los supuestos se procedió de la siguiente manera:

- **Normalidad Multivariada** se utilizó el test de Mardia.
- **Igualdad de Matrices de Varianza-Covarianza** se comprobó mediante la prueba de Box

Los supuestos de **Linealidad y Multicolinealidad y Singularidad** no fueron revisados, dado que para la construcción del modelo de análisis discriminante se utilizó el método por etapas o paso a paso, teniendo en cuenta el criterio de tolerancia para seleccionar las variables que son incluidas. De este modo, aquellas variables que presentan una correlación múltiple elevada con las restantes variables arrojarán una baja tolerancia y no serán consideradas de cara a la construcción de la función discriminante. El resultado final del modelo es el siguiente:

$$Z = 0,185 * (\text{Puntaje Química ICFES}) - 9,192$$

La capacidad de clasificación del modelo fue buena dando como resultado un error Tipo I de 29.4% y un error Tipo II de 25,0% para una efectividad de clasificación de 72,20%.

Resultados de la clasificación

		Fortaleza	Grupo de pertenencia		Total
			1	2	
Original	Recuento	1	24	10	34
		2	5	15	24
	%	1	70,6	29,4	100
		2	25,0	75,0	100
Validación cruzada ^a	Recuento	0	24	10	34
		1	5	15	24
	%	0	70,6	29,4	100
		1	25,0	75,0	100

6. Conclusiones

El procedimiento de Análisis Discriminante parece ser un instrumento idóneo para hacer una buena predicción del rendimiento académico como lo muestra el caso de aplicación desarrollado en donde el 72% de casos han sido bien clasificados. Con La propuesta del modelo de predicción de éxito académico de los estudiantes de primer semestre de la Universidad del Atlántico al lograr identificar las variables que influyen en el éxito académico esto se puede constituir en una herramienta al servicio del Bienestar Universitario para hacer seguimiento a los estudiantes que podrían presentar dificultades y brindarle todo el apoyo necesario para lograr aprobar el semestre. Se pretende iniciar el proyecto con los estudiantes de la facultad de Ingeniería y desarrollarlo durante el presente año.

Referencias

- Bacallao Gallestey J, Bacallao Guerra J, Parapar J. Roque M. (2004). Árboles de regresión y otras opciones metodológicas aplicadas a la predicción del rendimiento académico. Educación Médica Superior. Vol. 18, No 3
- Bonucci R. (1997). Qué variables predicen el rendimiento universitario de los estudiantes. Estudio de un caso: la facultad de Ingeniería en la U.L.A. Ponencia presentada en el III Simposio sobre políticas de admisión en la educación superior.
- Cuxart A, Martí M, Ferrer J. (1997). Algunos factores que inciden en el rendimiento y la evaluación en los alumnos de las pruebas de acceso a la universidad. Revista de Educación. Vol 314. 63-88.
- Dallas Johnson. Métodos multivariados aplicados al análisis de datos. Editorial Thomson Paraninfo.2000.
- García V, Alvarado J, Jiménez A. (2000). La predicción del rendimiento académico: regresión lineal versus regresión logística. Psicothema Vol. 12, No 2. 248-252
- Goberna, M.A., López M.A., Pastor J.T. (1987). La predicción del rendimiento como criterio para el ingreso en la universidad. Revista de Educación. Vol. 283. 235-248
- Gómez D, Oviedo R, Martínez E. (2011). Factores que influyen en el rendimiento académico del estudiante universitario. Educación y Humanidades.Vol 5. No 2. 90-97.

- Guzmán Martha. (2012). Modelos predictivos y explicativos del rendimiento académico universitario: caso de una institución privada en México. Tesis Doctoral. Universidad Complutense de Madrid.
- Jiménez Fernández C. (1987). Rendimiento académico en la Universidad a Distancia: un estudio empírico sobre su evolución y predicción. Revista de Educación. Vol 284. 317-347.
- Mileris Ricarda. (2010). Estimation of loan applicants default probability applying discriminant analysis and simple Bayesian classifier. Economics and Management. Vol. 33. No 15. 1078-1084.
- Moral José. (2006) .Predicción del rendimiento académico universitario. Perfiles Educativos. Vol 28. No 113. 38-63
- Pérez. César. Técnicas de análisis de datos con SPSS 15, ISBN: 978-848-322-601-8. Editorial Pearson, Prentice Hall, España, 2009
- Reyes J, Escobar C, Duarte J, Ramírez P. (2007). Una aplicación del modelo de regresión logística en la predicción del rendimiento estudiantil. Estudios Pedagógicos. Vol 33. No 2. 101-120.
- SPADIES 2012. Consultado en 01 de Diciembre de 2012 en: http://spadies.mineducacion.gov.co/spadies/consultas_predefinidas.html
- Wilson, R.L., Hardgrave, B.C. (1995). Predicting graduate student success in an MBA program: Regression versus classification. Educational and Psychological Measurement. Vol 55. 186-195.

Sobre los Autores

- **Adel Alfonso Mendoza Mendoza** es Magíster en Ingeniería Industrial .Profesor de Tiempo Completo de la Universidad del Atlántico adscrito al programa de Ingeniería Industrial. E-mail: adelmendoza@mail.uniatlantico.edu.co
- **Roberto José Herrera Acosta** es Magíster en Ciencias Estadísticas. .Profesor de Tiempo Completo de la Universidad del Atlántico adscrito al programa de Ingeniería Industrial. E-mail: robertoherrera@mail.uniatlantico.edu.co

Los puntos de vista expresados en este artículo no reflejan necesariamente la opinión de la Asociación Colombiana de Facultades de Ingeniería y de la International Federation of Engineering Education Societies

Copyright © 2013 Asociación Colombiana de Facultades de Ingeniería (ACOFI), International Federation of Engineering Education Societies (IFEES)