

Una explicación del rendimiento estudiantil universitario mediante modelos de regresión logística

Valera, Jorge;* Sinha, Surendra;**
Varela, José*** y Ponsot Balaguer, Ernesto****

Recibido: 12/01/2009 • Revisado: 05/03/2009
Aceptado: 15/03/2008

Resumen >>

Este trabajo estudia la relación entre el éxito académico de los estudiantes de ingeniería de la Universidad de los Andes, medido en función de su rendimiento académico, con algunas variables relacionadas con su entorno social y educativo desde el punto de vista de los modelos de regresión logística. Entre todas las variables consideradas, se determinó que sólo el promedio de bachillerato tiene un efecto significativo sobre el rendimiento académico en el primer semestre. Para explicar el rendimiento académico en el segundo semestre se encontró que sólo las variables promedio de calificaciones y sexo resultaron tener un efecto significativo. Los resultados mostraron que los estudiantes con promedios en bachillerato buenos o aceptables tienen mayores posibilidades de obtener un mejor rendimiento académico en el primer semestre, que aquellos con promedios deficientes en bachillerato. En cuanto al rendimiento académico en el segundo semestre, los datos indican que quienes tienen mayores posibilidades de obtener un mejor rendimiento académico en este semestre son los estudiantes de género femenino con promedios de notas buenos en el primer semestre.

Palabras clave: rendimiento académico, modelos logit, posibilidades proporcionales.

Abstract >>

An explication of student's university performance by logistic regression models

This study examines the relationship between academic success of engineering students from the University of los Andes, as measured by their performance, with some variables related to their educational and social environment from the point of view of the logistic regression models. Among all variables considered, it was determined that the average high school only has a significant effect on academic performance in the first semester. To explain academic performance in the second semester it was found that only the variables first semester average and sex were found to have a significant effect. The results showed that students with high school averages in good or fair are more likely to achieve better academic performance in the first semester, those with poor averages in high school. With regard to academic performance in the second semester, the data suggest that those who are more likely to achieve better academic performance are female students with good grade point averages in the first semester.

Key words: Academic performance, logit models, proportional possibilities.

*Ingeniero de Sistemas. Profesor en la Universidad Nacional Experimental del Táchira. jorgevalera39@hotmail.com

**B.Sc.(Agr.), (Bihar University, India) MSc y Doctor of Philosophy (PhD) in Statistical Genetics (Oregon State University, Corvallis, U.S.A.). Profesor Titular en la Universidad de Los Andes, Facultad de Ciencias Económicas y Sociales (FACES).sinha32@yahoo.com; <http://mipagina.cantv.net/ssinha/>

***Licenciado en Matemáticas. M.Sc. en Matemáticas. Profesor Asistente en el Instituto Universitario de Tecnológica de Cumaná. leovarela01@hotmail.com,

****Ingeniero de Sistemas. M. Sc. en Estadística Aplicada Profesor Asociado en la Universidad de Los Andes, Facultad de Ciencias Económicas y Sociales (FACES). ernesto@ula.ve; <http://webdelprofesor.ula.ve/economia/ernesto>

1. Introducción

Desde hace tiempo, el rendimiento estudiantil en las universidades ha sido un tema de interés para muchos investigadores. Conocer los principales factores que influyen sobre el rendimiento estudiantil y la manera como se produce esta influencia permitiría establecer políticas dirigidas a elevar el rendimiento de los jóvenes que ingresan a las universidades, lo cual traería como consecuencia profesionales mejor preparados, mayor oferta de cupos universitarios y en general un mayor desarrollo económico y social. Como Ferreres, González y Rodríguez (2006) expresan: “Todas las organizaciones internacionales –IEA (Asociación Internacional para el Rendimiento Académico), la OCDE y el Banco Mundial– comparten el convencimiento explícito de que la ‘mejor’ educación puede ser medida y que además se traduce directamente en una mayor productividad económica y social”. (p. 405)

Según Georgescu, Martínez, Osta y Rivas (1995) “bajos índices de aprobados en los primeros semestres y una alta deserción entre otros factores fue lo que hizo que el Consejo de la Facultad de Ingeniería de la Universidad de Carabobo en 1994 implementara la Prueba PAI (Prueba de Admisión de la Facultad de Ingeniería) en el proceso de admisión” (p. 2).

Por otra parte, cuando una persona ingresa al sistema de educación superior se da cuenta inmediatamente de que el método de estudio empleado para adquirir conocimiento en bachillerato es insuficiente. En educación media y diversificada el alumno se circunscribe a estudiar solo lo que el profesor imparte en sus clases, y la lectura, mas allá de los contenidos dados en el salón de clase y el auto-aprendizaje son prácticas casi nulas en la etapa de liceísta.

Al empezar los estudios universitarios, los estudiantes observan que para alcanzar el éxito es obligatorio no conformarse con los contenidos facilitados en el aula de clase y deben realizar trabajo de investigación. Lamentablemente, como Morles (1999) expresa “según los propios educadores, la inmensa mayoría de los niños y jóvenes venezolanos no ha leído un libro completo, una revista o una nota periodística, a pesar de que, supuestamente, saben hacerlo” (p. 67)

En este contexto, los conocimientos adquiridos en el bachillerato, el entorno económico-social y el cambio en la metodología de aprendizaje que se debe asumir en el primer semestre universitario, son entre otros, factores que ejercen gran influencia en la tasa de aprobación en los primeros dos semestres de vida universitaria y por ende en el rendimiento académico al comienzo de los estudios universitarios.

2. Planteamiento del problema

El bajo rendimiento estudiantil en los primeros semestres universitarios es un problema que deben enfrentar las universidades en la actualidad. Semestre a semestre se observa un alto porcentaje de reprobados en las asignaturas de los semestres iniciales, lo que trae como consecuencia que se produzca una “barrera académica” que impide que la oferta de cupos sea la deseada, ya que no se puede en el corto plazo incrementar la planta profesoral para estas asignaturas, ni habilitar nuevos espacios físicos en las universidades. Desde este punto de vista, para incrementar la oferta en las universidades es necesario que se lleven a cabo políticas que permitan mejorar significativamente el rendimiento estudiantil de

los universitarios en general, pero sobre todo de quienes cursan los primeros semestres de la carrera.

La implementación de las políticas adecuadas para combatir este problema requiere de un conocimiento científico de los factores que contribuyen a que se produzca el bajo rendimiento estudiantil. En carreras como ingeniería se observa bajo rendimiento académico inicial, y lamentablemente casi no existen programas de orientación que instruyan al estudiantado sobre cómo adaptarse al cambio y facilitar su adaptación al nuevo sistema. En la práctica, desde que un estudiante ingresa a la universidad, comienza un proceso de auto aprendizaje sobre el nuevo régimen de estudios, diferente para cada uno. En buena medida el éxito depende de las características particulares y no de políticas orientadoras generales. En este sentido, se hace necesario la construcción de mecanismos que permitan comprobar qué parte del éxito del estudiante se debe a su capacidad de adaptación al nuevo sistema, así como investigar los factores que mejor contribuyen a este proceso de adaptación, medidos en función de su capacidad no solo de aprobar asignaturas, sino aprobarlas con buenas calificaciones.

Este trabajo espera contribuir a esta tarea, especialmente arrojando luces sobre el rendimiento estudiantil, que puedan orientar decisiones de gerencia académica, las cuales propendan a mitigar los efectos del cambio en los alumnos de nuevo ingreso.

3. Marco teórico

El éxito académico de un estudiante universitario depende de muchos factores, y según González (1988), para analizar el bajo

rendimiento académico la variable Rendimiento Estudiantil tiene gran importancia. Por otra parte, González (1989) expresa "...el rendimiento en los dos primeros semestres de estudios universitarios es un buen predictor del rendimiento posterior, el factor aptitudinal es el que tiene mayor influencia en el rendimiento estudiantil y tanto la preparación que tiene el estudiante al entrar a la universidad como el factor socioeconómico, influyen directamente sobre el factor aptitudinal e indirectamente (a través de éste), sobre el rendimiento estudiantil" (p. 1).

Por otra parte, Arredondo y Zapata indican "Los investigadores en su afán por determinar los factores que influyen en el rendimiento académico, coinciden en señalar que éste depende de diversos aspectos, tales como: la personalidad, la edad, el sexo, los padres, la actitud, la aptitud y los métodos de enseñanza, entre otros".

Esto parece indicar que el éxito universitario está ligado al rendimiento académico en los primeros dos semestres, el cual en cierta medida depende del aprendizaje adquirido en el bachillerato, el entorno socioeconómico, y su capacidad de adaptación al nuevo régimen de estudios, esta última relacionada con el desarrollo de capacidades para realizar investigaciones bibliográficas, redacción de informes, capacidad de interacción académica con sus pares, y a la nueva relación de aprendizaje alumno-profesor, que a nivel universitario es menos personalizada.

Cuando se está cursando el segundo semestre universitario, el éxito académico que un estudiante pueda obtener puede aún depender del conocimiento adquirido durante el bachillerato, de su condición socioeconómica,

que en un lapso tan breve, se asume cambia poco, y del aprendizaje que sobre el sistema haya adquirido en el primer semestre. Así, esta investigación busca ajustar modelos de regresión logística para estudiar la relación entre estos y otros factores de interés, con el éxito académico de un estudiante de la Facultad de Ingeniería (FI), de la Universidad de los Andes (ULA), en su primer y segundo semestre de estudios, los cuales son definitorios del rendimiento estudiantil en los semestres superiores (Varela, Sinha, Ponsot y Valera, 2008).

4. Los modelos de regresión logística

Los modelos de regresión logística son modelos estadísticos que se utilizan ampliamente cuando se desea describir la relación existente entre una variable respuesta de tipo categórico y un conjunto de variables explicativas categóricas o continuas. Estos modelos son un caso particular de un tipo de modelos estadísticos más general conocido como Modelos Lineales Generalizados (MLG).

En todo MLG se pueden distinguir tres componentes:

Componente aleatorio. Identifica a la variable respuesta y asume para ésta una distribución de probabilidad. Cuando la respuesta es binaria la distribución de probabilidad que por lo general se asume es la Bernoulli, y en este caso los valores de la respuesta se clasifican como éxito o fracaso.

Componente sistemático. Es la parte del modelo relacionada con las variables explicativas. Es la combinación lineal ($X\beta$) de las variables explicativas X a utilizar del lado derecho de la ecuación que define al modelo. A esta combinación lineal se le conoce como el *predictor*

lineal, y a β como el vector de parámetros del modelo.

La función de enlace (función *link*). La función de enlace $g(\cdot)$ es una transformación uno a uno, continua y diferenciable de la media $\mu = E(Y)$ de la variable respuesta Y . El nombre de función de enlace proviene del hecho que la función $g(\cdot)$ enlaza los componentes aleatorio y sistemático del modelo; esto es, $g(\mu) = \alpha + \beta_1 X_1 + \dots + \beta_k X_k$.

En modelos de regresión logística la variable respuesta puede ser de naturaleza dicotómica, politómica nominal o bien politómica ordinal, siendo de éste último tipo, la variable respuesta que se estudia en esta investigación.

Denotemos con j el número de categorías posibles para la variable respuesta Y con probabilidades de ocurrencia de cada categoría dadas por $\{\pi_1, \pi_2, \dots, \pi_j\}$, las cuales satisfacen.

$\sum_{k=1}^j \pi_k = 1$ Cuando el experimento aleatorio para observar Y se realiza n veces de manera independiente, la distribución del número de éxitos en cada categoría k es multinomial. Cuando las categorías de la variable respuesta son de tipo ordinal, un modelo que se emplea con frecuencia es el *logit* acumulativo para respuesta ordinal. Este modelo utiliza el ordenamiento de la variable respuesta Y para obtener resultados fáciles de interpretar. En este modelo los *logit* acumulativos se definen en función de la probabilidad acumulativa de la variable Y , dada por $P(Y \leq k) = \pi_1 + \dots + \pi_k$ para $k = 1, \dots, j-1$ y se expresan como en (1).

$$\begin{aligned}
 (1) \quad \text{logit} [P(Y \leq k)] &= \log \left[\frac{P(Y \leq k)}{1 - P(Y \leq k)} \right] \\
 &= \log \left[\frac{\pi_1 + \dots + \pi_k}{1 - (\pi_1 + \dots + \pi_k)} \right] \\
 &= \log \left[\frac{\pi_1 + \dots + \pi_k}{\pi_{k+1} + \dots + \pi_j} \right]
 \end{aligned}$$

para $k = 1, \dots, j-1$. En el caso particular en que la respuesta posea tres categorías de respuesta, los *logits* son:

(2)

$$\text{logit} [P(Y \leq 1)] = \log [\pi_1 / (\pi_2 + \pi_3)]$$

$$\text{logit} [P(Y \leq 2)] = \log [(\pi_1 + \pi_2) / \pi_3]$$

5. El modelo logit acumulativo con posibilidad proporcional

El modelo para el k -ésimo *logit* acumulativo es similar a un modelo de regresión logística al combinar las categorías de la respuesta en dos grupos: uno con las categorías de la 1 a la k y el otro con las categorías de la $k+1$ a la j . Por ejemplo, si solo se tuviese una variable explicativa x , el modelo para el k -ésimo *logit* sería como en (3).

$$\text{logit} [P(Y \leq k)] = \alpha_k + \beta x, k = 1, \dots, j-1$$

donde el parámetro β describe el efecto de x sobre el logaritmo de la posibilidad de la respuesta en la categoría k o inferior. Es importante destacar que en este modelo el coeficiente β no tiene subíndice dado que el modelo de posibilidad proporcional asume que el efecto de x es el mismo para todos los $j-1$ *logits* acumulativos. Su ventaja cuando ajusta bien a los datos es su sencillez, pues solo requiere de un parámetro para describir el efecto de la variable x , en lugar de $j-1$ parámetros como sería el caso del modelo *logit* para respuesta nominal (Agresti, 2007).

En cuanto a los resultados producidos por este modelo, éstos son similares a los del modelo de regresión logística para una

respuesta dicotómica, solo que se utilizan las probabilidades acumulativas de la variable respuesta y su complemento para obtener los Cocientes de Disparidad (CD), conocidos también como “*odds ratio*”. Para comparar el efecto de dos valores x_1, x_2 de X sobre la respuesta se calcula

$$CD = \frac{P(Y \leq k | X = x_2) / P(Y > k | X = x_2)}{P(Y \leq k | X = x_1) / P(Y > k | X = x_1)} \quad (4)$$

Al obtener el logaritmo de este cociente, se obtiene la diferencia entre los *logits* acumulativos evaluados en los valores x_1, x_2 de X ; esto es,

$$\log(CD) = \log \left(\frac{P(Y \leq k | X = x_2) / P(Y > k | X = x_2)}{P(Y \leq k | X = x_1) / P(Y > k | X = x_1)} \right)$$

$$\log(P(Y \leq k | X = x_2) / P(Y > k | X = x_2))$$

$$\log(P(Y \leq k | X = x_1) / P(Y > k | X = x_1))$$

$$\text{logit}(P(Y \leq k | X = x_2)) - \text{logit}(P(Y \leq k | X = x_1))$$

$$\alpha_k + \beta x_2 - (\alpha_k + \beta x_1) = \beta (x_2 - x_1) \quad (5)$$

Luego, la diferencia de *logits* es proporcional a la diferencia $x_2 - x_1$, y en consecuencia el CD está dado por $e^{\beta(x_2 - x_1)}$. Es importante señalar que en este modelo, el valor del CD es independiente del valor k que define los *logits* acumulativos. A esta propiedad se le conoce como “el supuesto de posibilidad proporcional”, el cual siempre debe comprobarse para determinar su validez. Cuando el orden en que se acumulan los *logits* es el contrario, los resultados son los mismos pero con el signo contrario para β .

En general, cuando la respuesta es poltómica ordinal y se postula un modelo de posibilidad proporcional, se lleva a cabo la Prueba de Puntuación (*Score Test*) para verificar su sostenibilidad. El *Score Test*, es una prueba chi-cuadrado donde la hipótesis nula es

$H_0: \beta_k = \beta$ para todo k , siendo β el vector de coeficientes de las variables explicativas incluidas en el modelo. Seguidamente se llevan a cabo las pruebas de ajuste del modelo, conocidas como la prueba chi-cuadrado de Pearson y *deviance* G^2 , que comparan la frecuencia observada y esperada en cada categoría de la respuesta, para cada posible combinación de las variables explicativas. Una vez que el modelo ajustado haya pasado estas pruebas, se procede a observar la significación estadística de las variables explicativas incluidas en él a través del estadístico de Wald (Rodríguez, s/f).

6. Metodología

6.1 Fuente de datos

La base de datos de donde se llevó a cabo la selección de las variables y la muestra de datos utilizada en este estudio, se obtuvo gracias al Ing. Gustavo Briceño Torres, Coordinador de Desarrollo de la Dirección de Servicios de Información Administrativa de la ULA. Entre las actividades que lleva a cabo esta Dirección está la de recopilar los datos provenientes del Sistema Integrado de Registros Estudiantiles (ULA_SIRE), y del Sistema Central de Registros Estudiantiles (ULA_SCRE). El primer sistema acompaña las inscripciones estudiantiles en la ULA y el segundo es el responsable de recopilar la información relativa al ingreso y egreso de los alumnos de la universidad.

6.2 Población y muestra

La población de interés está compuesta por 3962 estudiantes de la Facultad de Ingeniería de la Universidad de los Andes, quienes comenzaron sus estudios entre los años 2003 al 2006. De esta población se tomó una muestra de 458 estudiantes, seleccionados

con base en la información disponible en la base de datos consolidada del ULA_SIRE y el ULA_SCRE. Ambos sistemas contienen información sobre alumnos, carreras, *pensa* de estudios, asignaturas cursadas en distintos períodos, calificaciones obtenidas, información demográfica y socioeconómica del alumno, información del plantel en que cursara estudios secundarios e información de ingreso, egreso y cambio de carrera.

El ULA_SCRE entró en operación a partir del año 2000, mientras que el ULA_SIRE lo hizo en el año 2002. Ambos sistemas comenzaron con un proceso de migración de datos de los sistemas anteriores, lo que proporciona información de alumnos cuyo ingreso se remonta al año 1974 y llega hasta el presente. En conversaciones sostenidas con quienes proporcionaron la información, se concluyó que la máxima confiabilidad se obtendría en aquellos registros de alumnos cuyo ingreso haya ocurrido desde el año 2003. Esto introduce una primera reducción en los datos para el estudio. Adicionalmente, no se consideraron las asignaturas retiradas legalmente por el alumno, ni aquellas cuya calificación fue asentada estando el alumno incurrido en medidas de rendimiento estudiantil. Por otra parte, luego de una depuración mayor que incluyó la selección de aquellos alumnos con calificaciones de bachillerato válidas, así como con información sobre la región a la que pertenece el instituto de educación media en que egresaron, quedaron los 458 registros de estudiantes empleados.

Finalmente, es necesario acotar que en vista de que el proceso de depuración y filtrado de los datos responde a criterios fortuitos, sin sesgo de parte del investigador, no hay razones para dudar de la representatividad (e incluso, aleatoriedad) de la muestra con la cual se cuenta para el estudio.

6.3 Especificación y codificación de las variables

De las variables disponibles en la base de datos, las que fueron consideradas de interés para esta investigación se describen en la Tabla N° 1. La variable Rendimiento Académico en el Semestre (RAS), con valores entre 0% y 100%, se expresa en (6).

$$RAS = \frac{\text{Número de materias aprobadas en el semestre}}{\text{Número de materias inscritas en el semestre}} \times 100$$

6.4 Estudio descriptivo de las variables y su interrelación

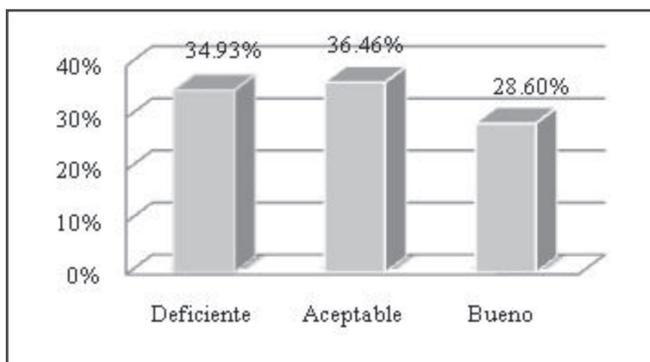
La muestra está compuesta por 62,66% de estudiantes del sexo femenino, el porcentaje de estudiantes con notas de bachillerato clasificadas como aceptables, buenas y malas son similares (Figura 1), y el 37% de las madres poseen nivel de instrucción universitario. Además, un 88% de los estudiantes de la muestra obtuvo su título de bachiller en la región andina, e igual porcentaje obtuvo el título de bachiller en Ciencias.

Tabla 1. >>>
Codificación de las variables de interés en el estudio

	Variable	Valores originales	Codificación
Independientes	Promedio de notas de bachillerato.	10 a 20 puntos	10 – 14: Deficiente 15–16: Aceptable 17 – 20: Bueno
	Sexo	Masculino Femenino	M: Masculino F: Femenino
	Nivel instrucción de la madre	Universitaria Secundaria completa Secundaria incompleta Analfabeta	Universitaria No universitaria
	Región donde cursó bachillerato	Los Andes Otra región	Andes No Andes
	Título obtenido de educación media	Bachiller en ciencias Otras menciones	Ciencias Otro
	Edad	De 18 o menos años 19-20 años De 21 o más años	18 y menos 19-20 21 y más
	Promedio de notas en el primer semestre	0 – 20 puntos	< 10: Deficiente 10 – 13: Regular 14 – 16: Aceptable 17 – 20: Bueno
Promedio de notas en el segundo semestre	0 – 20 puntos	< 10: Deficiente 10 – 13: Regular 14 – 16: Aceptable 17 – 20: Bueno	
Dependientes	Rendimiento académico en el primer semestre	0 – 100%	< 50: Deficiente 50 – <75: Aceptable ≥ 75: Bueno
	Rendimiento académico en el segundo semestre	0 – 100%	< 50: Deficiente 50 – <75: Aceptable ≥ 75: Bueno

Fig. 1. >>

Distribución estudiantil según el promedio de notas de bachillerato



En lo que respecta a la edad, el 43,89% de los estudiantes de la muestra comenzaron sus estudios de ingeniería a la edad de 18 años o menos, el 30,79% de 19 o 20 años, y el resto de 21 o más. En la Tabla N° 2 se muestran algunas estadísticas descriptivas de interés para las variables Edad, Promedio de notas bachillerato, Promedio de notas en el primer semestre, Rendimiento académico en el primer semestre y Rendimiento académico en el segundo semestre.

En la tabla 2 se observa: (1) Existen correlaciones muy bajas entre edad y las demás variables, lo cual parece sugerir que la edad no es una buena predictora del rendimiento académico. (2) Existen altas correlaciones positivas entre el promedio de notas en el semestre y el rendimiento académico en el mismo Semestre. Esto implica que buenos promedios implican buenos rendimientos. Así, solo se ajustarán modelos para explicar las variables rendimiento en el primer semestre y rendimiento en el segundo semestre. (3) Existen correlaciones intermedias entre el promedio de notas de bachillerato con el rendimiento académico en el primer y segundo semestre, y con los promedios de notas en los semestres.

7. Construcción de los modelos de regresión logística

Dada la alta correlación entre el promedio de notas en el semestre con el correspondiente Rendimiento académico en el mismo semestre, se decidió obtener modelos de regresión logística que expliquen las variables Rendimiento académico en el primer semestre y Rendimiento académico en el segundo semestre, en función de las variables independientes disponibles. Para ello se aplicó el método de selección de las variables independientes “*stepwise*”, que consiste en agregar al modelo una a una las variables independientes y observar su efecto, ya que como menciona Canavos (1988): “La inclusión de una variable de predicción en un modelo de regresión no implica, en forma necesaria, que tenga un efecto sustancial sobre la respuesta dada; es decir, cuando un investigador identifica un conjunto de variables de predicción, esto indica el potencial de las variables para explicar la variación en la respuesta. Queda por comprobarse si algunas realmente lo hacen” (p. 513). Este método retiene aquellas variables independientes cuyo efecto sobre la respuesta resulta estadísticamente significativo.

7.1 Selección del mejor modelo de ajuste del rendimiento académico

Se aplicaron modelos de regresión logística para explicar el Rendimiento académico en el primer semestre y el Rendimiento académico en el segundo semestre. En cada caso, se ajustó un modelo de posibilidad proporcional (McCullagh, 1980), el cual permite la modelación de los *logits* acumulativos de la variable respuesta cuando ésta es de naturaleza ordinal, cual es el caso.

7.2 Modelado del Rendimiento académico del primer semestre (RAPS)

Los resultados obtenidos al aplicar el método *stepwise* determinaron que sólo la variable Promedio de notas de bachillerato (Prom_Bachi) posee un efecto significativo sobre la respuesta RAPS. Por otra parte, el valor chi-cuadrado del *Score Test* con 4 grados de libertad es 1,2868, con un valor_p = 0,5255 indicando que

la hipótesis “ H_0 : Es apropiado emplear el modelo de posibilidad proporcional” no se rechaza. Los estadísticos Devianza (valor_p = 0,5327) y chi-cuadrado de Pearson (valor_p = 0,5361) revelan un ajuste adecuado del modelo.

En la Tabla N° 3 se presentan los valores estimados de los dos interceptos involucrados en el modelo y de los coeficientes asociados a la variable Prom_Bachi, los cuales son todos altamente significativos. Igualmente se presenta la información relativa a los estimadores de cocientes de disparidad. Se puede decir que un estudiante con promedio de notas en bachillerato considerado bueno tiene 3,44 veces más posibilidad de obtener un rendimiento académico en su primer semestre de ingeniería bueno, en lugar de aceptable o deficiente, o bien de obtener un rendimiento académico en su primer semestre de ingeniería bueno o aceptable, en lugar de deficiente, que un estudiante cuyo promedio en el bachillerato fue deficiente.

Tabla 3. >>

Resultados para el Modelo Logit Acumulativo del Rendimiento académico en el primer semestre.

Análisis del estimador de máxima verosimilitud						
Parámetro		DF	Estimador	Error estándar	Chi-cuadrado de Wald	Pr > ChiSq
Intercepto	Bueno	1	-0.7267	0.1232	34.7873	<.0001
Intercepto	Aceptable	1	0.3282	0.1205	7.4208	0.0064
Prom_Bachi	Bueno	1	1.2343	0.1831	45.4278	<.0001
Prom_Bachi	Aceptable	1	0.5701	0.1636	12.1366	0.0005
Estimadores de cocientes de disparidad						
Efecto		Estimador puntual		Límites de confianza de Wald del 95%		
Prom_Bachi Bueno vs Deficiente		3.436		2.400	4.919	
Prom_Bachi Aceptable vs Deficiente		1.768		1.283	2.437	

De igual manera, un estudiante con promedio de notas en bachillerato aceptable tiene 1,77 veces más posibilidad de obtener un rendimiento académico en su primer semestre de ingeniería bueno, en lugar de aceptable o deficiente, o bien de obtener un rendimiento académico en su primer semestre de ingeniería bueno o aceptable, en lugar de deficiente, que un estudiante cuyo promedio en el bachillerato fue deficiente.

A partir de la ecuación (3) y de los resultados anteriores obtenemos que el modelo *logit* acumulativo está dado por la expresión:

$$\text{logit} [P (RAPS \leq k)] = \alpha_k + \beta_1 x_1 + \beta_2 x_2 = 1, 2 \quad (7)$$

donde $\alpha_1 = -0.73$, $\alpha_2 = 0.33$, $\beta_1 = 1.23$, parámetro que representa el nivel “Bueno” de la variable Prom_Bachi y $\beta_2 = 0.57$, parámetro que representa el nivel “Aceptable” de la variable Prom_Bachi. El nivel de referencia contra el cual se realiza el estudio de posibilidades es la categoría “Deficiente” de la variable Prom_Bachi.

En este caso las probabilidades acumulativas están dadas por:

(8)

$$P (RAPS \leq 1) = P (\text{Rend. Bueno})$$

$$P (RAPS \leq 2) = P (\text{Rend. Bueno}) + P (\text{Rend. Aceptable})$$

Respecto a las variables explicativas x_1, x_2 en el modelo, se definen como:

(9)

$$x_1 = \begin{cases} 1 & \text{si Promedio de Bachillerato} = \text{Bueno} \\ 0 & \text{en cualquier otro caso} \end{cases}$$

$$x_2 = \begin{cases} 1 & \text{si Promedio de Bachillerato} = \text{Aceptable} \\ 0 & \text{en cualquier otro caso} \end{cases}$$

7.3 Modelado del rendimiento académico del segundo semestre (RASS)

Similar al caso para el modelado del rendimiento académico en el primer semestre, se aplicó el método *stepwise*. Los resultados obtenidos indican que las variables: Calificación promedio en el primer semestre (Prom_Sem1) y Sexo fueron las únicas que resultaron tener influencia significativa sobre la variable RASS.

En lo que respecta al valor de la prueba de puntuación, el valor chi-cuadrado con 4 grados de libertad que se obtuvo es 1,8262, con un valor_p = 0,7677, por lo que la hipótesis “H₀: Es apropiado emplear el modelo de posibilidad proporcional” no se rechaza. Por otra parte, se obtuvieron valores de *deviance* = 1,4796 y chi-cuadrado de Pearson = 1,3422, con valores_p de 0,1397 y 0,2011 respectivamente. Estos últimos resultados indican que el modelo ajusta a los datos.

Los resultados de la Tabla 4 muestran que todos los coeficientes del modelo son altamente significativos. De los valores obtenidos para los CD, se puede decir que un estudiante con un Prom_Sem1: bueno, aceptable o regular tiene 10,7 y 4,5 veces más posibilidad respectivamente de obtener en su segundo semestre de estudios de ingeniería, un rendimiento académico bueno en lugar de uno aceptable, regular o deficiente; o bien de obtener en su segundo semestre un rendimiento académico bueno o aceptable en lugar de uno regular o deficiente; o de obtener en su segundo semestre un rendimiento académico bueno, aceptable o regular en lugar de uno deficiente, que un estudiante con un Prom_Sem1 deficiente.

En lo que respecta a la variable sexo, se aprecia que los estudiantes del sexo femenino

tienen casi el doble de posibilidad (1/0,52) de obtener en el segundo semestre de ingeniería un rendimiento académico bueno en lugar de aceptable o deficiente; o bien de obtener en el

segundo semestre de ingeniería un rendimiento académico bueno o aceptable en lugar de uno deficiente, que sus pares masculinos.

Tabla 4. >>>
Resultados para el Modelo Logit Acumulativo del Rendimiento académico en el segundo semestre

Análisis del estimador de máxima verosimilitud						
Parámetro		DF	Estimador	Error estándar	Chi-cuadrado de Wald	Pr > ChiSq
Intercept	Bueno	1	-0.5437	0.2168	6.2901	0.0121
Intercept	Aceptable	1	0.5984	0.2158	7.6885	0.0056
Prom_Sem1	Bueno	1	2.3022	0.5901	15.2223	<.0001
Prom_Sem1	Aceptable	1	1.9345	0.3861	25.1068	<.0001
Prom_Sem1	Regular	1	1.4920	0.2796	28.4650	<.0001
Sexo	M	1	-0.6585	0.2368	7.7301	0.0054
Estimadores de cocientes de disparidad;						
Efecto			Estimador del punto	Límites de confianza de Wald del 95%		
Prom_Sem1 Bueno vs Deficiente			9.997	3.145	31.778	
Prom_Sem1 Aceptable vs Deficiente			6.921	3.247	14.750	
Prom_Sem1 Regular vs Deficiente			4.446	2.570	7.691	
Sexo M vs F			0.518	0.325	0.823	

También de la Tabla 4 se obtiene la expresión del modelo *logit* acumulativo que ha sido ajustado, como en (10).

$$\text{logit} [P(\text{RASS} \leq k)] = \alpha_k + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$$

para $k = 1, 2$ (10)

En este caso, los valores estimados de los parámetros son

$$\alpha_1 = -0,5437, \alpha_2 = 0,5984, \beta_1 = 2,3022, \beta_2 = 1,9345, \beta_3 = 1,4920, \text{ y } \beta_4 = -0,6585$$

Al igual que antes, las probabilidades acumulativas para este modelo se definen como en (11).

$$P(\text{RAPS} \leq 1) = P(\text{Rend. Bueno})$$

$$P(\text{RAPS} \leq 2) = P(\text{Rend. Bueno}) + P(\text{Rend. Aceptable})$$

Las variables se definen como en (12).

$$x_1 = \begin{cases} 1 & \text{si Promedio del Primer Semestre} = \text{Bueno} \\ 0 & \text{en cualquier otro caso} \end{cases}$$

$$x_2 = \begin{cases} 1 & \text{si Promedio del Primer Semestre} = \text{Aceptable} \\ 0 & \text{en cualquier otro caso} \end{cases}$$

$$x_3 = \begin{cases} 1 & \text{si Promedio del Primer Semestre} = \text{Regular} \\ 0 & \text{en cualquier otro caso} \end{cases}$$

$$x_4 = \begin{cases} 1 & \text{si Sexo del Estudiante} = \text{Masculino} \\ 0 & \text{en cualquier otro caso} \end{cases}$$

Conclusiones

El éxito académico de un estudiante universitario está asociado con el rendimiento calculado como la proporción de asignaturas aprobadas en relación con el total de ellas inscritas, o bien en función de su promedio en el semestre. Del análisis de los datos se desprende que existe una alta correlación positiva entre las variables Rendimiento académico y el Promedio de notas en el mismo semestre, por lo que se decidió considerar sólo el efecto de algunas variables independientes de interés sobre la variable Rendimiento académico.

De las variables disponibles en la base de datos, se consideraron aquellas independientes que podían tener efecto sobre la respuesta de interés, en un modelo de regresión logística cuyo objetivo fuese explicar el Rendimiento académico. Los resultados obtenidos determinaron que de todas las variables consideradas, sólo el Promedio de bachillerato resultó tener un efecto significativo sobre el Rendimiento académico en el primer semestre. Para explicar el Rendimiento académico en el segundo semestre se encontró que las variables Promedio de calificaciones y Sexo fueron las que resultaron tener un efecto significativo.

Los resultados también mostraron que los estudiantes con promedios en bachillerato buenos o aceptables tienen mayores posibilidades de obtener un mejor Rendimiento académico en el primer semestre, que aquellos con promedios deficientes en bachillerato. En cuanto al Rendimiento académico en el segundo semestre, los datos indican que quienes tienen mayores posibilidades de obtener un mejor rendimiento académico en este semestre son los estudiantes de género femenino con promedios de notas buenos en el primer semestre.

Podría parecer evidente que a mayor rendimiento en los estudios secundarios se espere mayor rendimiento en los estudios universitarios, sin embargo, no es frecuente encontrar evidencia científica que apoye esta tesis. Por lo general, la evidencia es descriptiva exclusivamente. Para el caso, el modelo de regresión logística empleado proporciona evidencia de carácter inferencial, con la solidez suficiente para apoyar la toma de decisiones sobre la materia.

En cuanto a la gerencia académica universitaria, es paradójico encontrar que muchas de las universidades nacionales en el país (incluida la ULA), hayan decidido obviar los resultados en educación media de sus aspirantes, a la hora de establecer políticas de ingreso. Evidentemente, vistas las conclusiones de esta investigación, es claro que tal decisión no es acertada.

Finalmente, sobre las razones del éxito o del fracaso de un alumno en su primer semestre de universidad, los resultados descartan la mayoría de las variables analizadas y se concentran en el promedio de notas de la secundaria como única variable explicativa. Como política académica

universitaria, resulta consecuentemente importante establecer vínculos estrechos con los planteles de educación media, procurando elevar las calificaciones que obtienen los

alumnos en esta etapa de sus vidas, como medida que propenda a elevar sus posibilidades de éxito en la universidad.

Bibliografía >>

- Arredondo I. y Zapata R. (s/f). *El textocentrismo* (Documento en línea) 1-11. Disponible: www.monografias.com (Consulta 2008, febrero 14).
- Agresti, A. (2007) *An Introduction to Categorical Data Analysis* (2nd. edition). New Jersey USA: Wiley.
- Canavos, G. (1988). *Probabilidad y estadística. Aplicaciones y métodos*. México, Editorial Mc Graw-Hill.
- Ferreres Pavía, V., González Soto, A. y Rodríguez Espinar, S. (2006). *Evaluación para la mejora de los centros docentes*. Barcelona: Praxis.
- Georgescu B., I., Martínez G., J., Osta T., K. y Rivas F., E. (1995). Prueba de Admisión de la Facultad de Ingeniería de la Universidad de Carabobo y su incidencia en el rendimiento estudiantil. *Órgano de divulgación científica y tecnológica de la Facultad de Ingeniería de la Universidad de Carabobo*, 10(1)
- González M., P. (1988). Indicadores Sintéticos del Rendimiento Estudiantil. *Revista Economía N° 2. Universidad de los Andes*. 10(1). 69-84.
- González M., P. (1989). Aplicación del Lisrel al Análisis del Rendimiento Estudiantil. *Revista Economía N° 4. Universidad de los Andes*. 10(1). 55-73.
- McCullagh, P. (1980). Regression models for ordinal data (with discussion). *J. R. Stat. Soc. A, B42*: 109-142.
- Morles, A. (1999). Los libros de texto y la enseñanza de la lectura en Venezuela. *Investigación y Postgrado*. Vol 14., N° 2, Octubre.
- Rodríguez Germán, (s/f). Generalized Linear Models (Documento en línea). <http://data.princeton.edu/wws509/notes/a2.pdf> (Consulta 2008, Marzo 10).
- Varela, L., Sinha, S., Ponsot, E. y Valera, J. (2008). Valor pronóstico del k-ésimo período académico en el rendimiento de los alumnos de la FACES-ULA. Sometido a la consideración de la Revista Actualidad Contable FACES.