

Un modelo de regresión logística del rendimiento en los estudios universitarios: Caso FACES-ULA

Ponsot B., Ernesto; Sinha, Surendra;
Varela, Leonardo; Valera, Jorge

Ponsot B., Ernesto

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>

Varela, Leonardo

Licenciado en Matemáticas.
M.Sc. en Matemáticas. Profesor Asistente
en el Instituto Universitario de Tecnológica
de Cumaná.
leovarela01@hotmail.com, l
eovarela01@gmail.com

Sinha, Surendra

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 en
la Universidad de Los Andes, Facultad de
Ciencias Económicas y Sociales (FACES).
sinha32@yahoo.com;
<http://mipagina.cantv.net/ssinha/>

Valera, Jorge

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

Recibido: 21-05-2008
Revisado: 19-11-2008
Aceptado: 30-01-2009

Se presenta un modelo de regresión logística útil para esclarecer relaciones de causalidad entre el promedio de calificaciones de bachillerato, el sector del plantel de procedencia, la edad al ingreso, el sexo y el turno del plantel de procedencia de un alumno, sobre el resultado académico en sus estudios universitarios. Se estudia una muestra de 287 estudiantes de la Facultad de Ciencias Económicas y Sociales (FACES) de la Universidad de Los Andes (ULA), con ingreso de los años 2003 a 2007. Se describen las variables y se prueba estadísticamente que: (1) Hay diferencias en el rendimiento estudiantil entre carreras de la FACES; (2) el rendimiento universitario puede ser explicado a partir del rendimiento en los estudios secundarios; (3) las restantes variables no tienen efecto sobre el rendimiento universitario, y (4) no hay diferencias entre los promedios de calificaciones secundarias de alumnos provenientes de planteles públicos y privados.

Palabras clave: Regresión logística, rendimiento en educación media, rendimiento estudiantil universitario.

RESUMEN

We present a Logistic regression model, useful to clarify causality relations among the average high school graduates grades, high school origin, age at the beginning of the university studies, gender and the high school shift on the university academic results of a student. A sample of 287 students of the Faculty of Economics and Social Sciences (FACES) of the University of Los Andes (ULA) is studied with admission between the years 2003 to 2007. The variables are described and tested statistically that: (1) There are differences in the student performance among careers of the FACES, (2): The university performance can be explained by his yield in high school studies, (3) The remaining variables do not have effect on the university performance and (4) There are no differences between the average high school grades of students from public and private sectors

Keywords: Logistic regression, student high school performance, student university performance.

ABSTRACT

1. Introducción

El estudio del rendimiento estudiantil en la universidad ha sido por muchos años labor constante de los investigadores del proceso educativo. Conocer las causas que puedan atribuirse al fracaso o al éxito de un alumno es relevante para la universidad y para el país en general, cuanto más si la referencia es un país en vías de desarrollo como Venezuela, cuyas universidades en su mayoría están bajo la administración del Estado y son de enseñanza prácticamente gratuita. Luego, maximizar las posibilidades de éxito de una persona en sus estudios universitarios es relevante, puesto que ellos implican considerables erogaciones al presupuesto nacional, lo cual además limita la oferta de plazas a disposición de los nuevos bachilleres que año tras año tocan a las puertas del subsistema.

Muchos de los análisis que han sido publicados sobre la materia son esencialmente descriptivos, centrados en la presentación de proporciones y cifras, sin mayor atención a la inferencia formal (ver por ejemplo Porto et al, 2001 y Di Gresia *et al.*, 2002). Aun cuando esto no significa que tales estudios carezcan de valor o rigor científico, sí señala una carencia importante: La estadística que se utiliza para argumentar no concluye en estimaciones en el sentido estricto de probar hipótesis, o construir intervalos de confianza con grados controlados de certidumbre. En nuestra región, sin embargo, en la pasada década, se condujeron estudios más completos como por ejemplo los trabajos de González (1989) y Garnica *et al.* (1991) que abordan el tema basados en modelos LISREL, y Análisis Discriminante, respectivamente.

Consciente de estas carencias, este trabajo propone un estudio del grado de causalidad entre algunas de las variables que comúnmente integran el pronóstico sobre rendimiento universitario pero, a diferencia de

otros enfoques, utiliza, además de la estadística descriptiva, modelos de regresión logística de posibilidades proporcionales. Estos modelos hacen posible la inferencia clásica y, mejor aún, hacen factible el estudio comparativo de posibilidades entre los distintos niveles de las variables intervinientes con todo el rigor científico necesario.

Aunque los modelos que aquí se proponen pueden ser aplicados en general, el trabajo se circunscribe a la Facultad de Ciencias Económicas y Sociales (FACES) de la ULA, como muestra de lo que podría lograrse si se emprendiese un análisis similar ampliado a todas las facultades de la universidad e incluso a todas las universidades nacionales.

2. Justificación

En Venezuela hace tiempo está planteada una disputa entre el Estado, en su condición de ductor y financista de la educación superior, y las universidades autónomas, como receptoras de los fondos y de las políticas estatales, en cuanto al rendimiento de los alumnos en las aulas y el ingreso de los jóvenes bachilleres a la universidad. El primero alega que las universidades, aun cuando forman parte del Estado, se han tornado cada vez más excluyentes en la admisión de alumnos provenientes de clases sociales humildes (Fuenmayor *et al.*, 2002; González *et al.*, 2002). Se acusa a las instituciones de privilegiar el rendimiento estudiantil a secas, sin considerar atenuantes sociales ni económicos. Las universidades, por su parte, asignan toda la responsabilidad al Estado en el rechazo de gran cantidad de aspirantes que año a año solicitan ingreso, y argumentan insuficiencia de recursos (profesores, planta física, bibliotecas, etc.) para atender a más alumnos e ineficacia de las instituciones de educación media (especialmente de aquellas conducidas por el Estado) para formar alumnos verdaderamente

capaces de hacer frente a los retos que les depara una carrera profesional.

De un tiempo a esta parte el Ejecutivo Nacional ha venido insistiendo en que las universidades autónomas (como es el caso de la ULA) deben abrir sus puertas y permitir el ingreso de nuevos alumnos sin anteponer pruebas de conocimientos que les filtren. En otras palabras, desde el Estado se presiona a las universidades para que la selección de nuevos alumnos salga completamente de sus manos y pase a ser responsabilidad exclusiva de los entes gubernamentales. Desde el nivel de gobierno se hará la selección, no en base a resultados de pruebas de admisión o similares, sino privilegiando directamente a los alumnos que provengan del sector de educación media estatal, tal vez considerando las calificaciones obtenidas, pero en todo caso dentro de cada subgrupo socioeconómico, y nunca sobre el conglomerado total.

Debido a la gran demanda que ha tenido la ULA, el ingreso a sus aulas de un nuevo alumno ha ido migrando hacia un sistema de pruebas de selección interna que, hoy por hoy, representa el 39% de los nuevos ingresos que se producen, según las cifras del Sistema Central de Registros Estudiantiles.

Atendiendo a la discusión planteada, y por considerar que ninguno de los dos enfoques en pugna ha mostrado análisis suficiente y con la profundidad necesaria, este trabajo se propone contribuir y discutir sobre la validez de las argumentaciones de ambos sectores mediante el desarrollo de modelos estadísticos que permitan responder interrogantes: ¿es posible explicar el rendimiento en la universidad a partir del rendimiento en el bachillerato?, ¿a partir del sector de procedencia del plantel?, ¿en función de la carrera que cursa el alumno?, ¿con variables como la edad, el sexo o el turno del plantel?, entre otras.

3. Hipótesis y objetivos

El interés de la investigación se centra en contrastar estadísticamente las siguientes hipótesis:

- a) El rendimiento y la eficiencia en la carrera universitaria se explican principalmente por la carrera que cursa el alumno, el promedio de calificaciones que obtuvo en sus estudios secundarios, el tipo (público o privado) de plantel de procedencia, el sexo, la edad al ingreso y el turno del plantel de procedencia, y
- b) Los promedios de calificaciones obtenidos por los alumnos en sus estudios secundarios, son estadísticamente diferentes dependiendo de si ellos provienen de planteles públicos o privados, o de si provienen de planteles diurnos u otros.

Atendiendo a estas hipótesis, se señalan a continuación los objetivos del estudio.

3.1. Objetivo general

Estudiar las relaciones de causalidad de las variables carrera que cursa (Carrera), promedio de educación media (PromBR), sector del plantel de procedencia (SectorP), edad al momento del ingreso (Edad), sexo del alumno (Sexo), turno del plantel de procedencia (Turno), con respecto al rendimiento de un alumno durante sus estudios universitarios en la FACES, ULA. El rendimiento será medido de forma independiente tanto por su promedio de calificaciones en la universidad (PromU), como por la variable que se llamará eficiencia (Efi), entendida como la razón entre el número de asignaturas aprobadas e inscritas por el alumno en su trayectoria académica.

3.2. Objetivos específicos

1. Analizar y describir la muestra empleada.
2. Establecer estadísticamente, mediante un modelo lineal generalizado del tipo regresión logística con probabilidades pro-

porcionales, de qué manera las variables Carrera, PromBR, SectorP, Edad, Sexo y Turno explican el resultado de los estudios universitarios en la FACES-ULA, medido con las variables PromU y Efi consideradas por separado.

3. Establecer estadísticamente, mediante un modelo lineal generalizado del tipo regresión logística multinomial, de qué manera las variables SectorP y Turno explican el promedio de calificaciones de los estudios secundarios de un alumno de la FACES-ULA, medido con la variable PromBR.

4. Análisis descriptivo de los datos

4.1. Variables

El cuadro 1 muestra cada una de estas variables con sus niveles y restricciones.

Como puede apreciarse en el cuadro 1, las variables de respuesta (PromU y Efi) son de naturaleza continua pero han sido objeto de una categorización para el estudio. Se tomarán en cuenta tres niveles de la respuesta en cada caso, de tal forma que el orden de categorización (3 a 1) refleje el aumento en la calificación promedio o la eficiencia. También se ha realizado este proceso para las variables explicativas PromBR y Edad.

Cuadro 1
Lista de variables seleccionadas para el estudio

Variable	Descripción	Rango original	Rótulo de categoría	Rango de categoría
PromU	Promedio en la universidad	0 - 20	3. Rep. [0,10) 2. Inf. [10,14) 1. Sup. [14,20]	[0, 10) [10, 14) [14, 20]
Efi	$\frac{N^{\circ}\text{Mat.Aprob.} \times 100}{N^{\circ}\text{Mat.Cursadas}}$	0% - 100%	3. Inf. [0,65)% 2. Med. [65,85)% 1. Sup. [85,100)%	[0%, 65) [65%, 85%) [85%, 100%)
Carrera	Carrera que cursa	- - - -	Administración Contaduría Economía Estadística	- - - -
PromBR	Promedio en Bachillerato	10 - 20	2. Inf. [10,14) 1. Sup. [14,20]	[10, 14) [14, 20]
SectorP	Sector del plantel	- -	2. Priv./Otro 1. Púb.	- -
Sexo	Sexo del alumno	- -	F (Femenino) M (Masculino)	- -
Edad	Edad aproximada al ingreso	0 - 48	3. Años 21 o más 2. Años 19 - 20 1. Años 17 - 18	21 o más 19 - 20 17 - 18
Turno	Turno del plantel	- -	D (Diurno) N u O (Otro)	- -

4.2. Cuadros de contingencia de las variables principales

Los cuadros de contingencia 2 a 6 contienen las frecuencias de las variables SectorP y PromBR versus las variables respuesta PromU y Efi, para cada carrera.

En el cuadro 2, entre otras cuestiones, se aprecia que el estudio cuenta con 95 alumnos de la Licenciatura en Administración de Empresas, 47,37% proveniente de planteles del sector privado y 52,63% de planteles del

sector público. 41,05% de los alumnos llega a la universidad con promedios bajos en sus estudios de bachillerato, de ellos, apenas 15,38% logra promedios elevados en la universidad y 35,90% logra un nivel elevado de eficiencia. Por otra parte, 58,95% de los alumnos llega a la universidad con promedios altos en sus estudios de bachillerato, y de ellos, 26,79% logra promedios elevados en la universidad y 42,86% logra un nivel elevado de eficiencia.

Cuadro 2
Contingencia para PromU y Efi: Administración

SectorP	PromBR	PromU			Efi			Total
		3. Rep. [0,10]	2. Inf. [10,14]	1. Sup. [14,20]	3. Inf. [0,65]%	2. Med. [65,85]%	1. Sup. [85,100]%	
2. Priv./Otro	2. Inf. [10,14]	9	8	4	8	4	9	21
		42,86	38,1	19,05	38,1	19,05	42,86	100
		69,23	42,11	30,77	61,54	36,36	42,86	46,67
	1. Sup. [14,20]	4	11	9	5	7	12	24
		16,67	45,83	37,5	20,83	29,17	50	100
		30,77	57,89	69,23	38,46	63,64	57,14	53,33
Total	4,21	11,58	9,47	5,26	7,37	12,63	25,26	
	13	19	13	13	11	21	45	
	28,89	42,22	28,89	28,89	24,44	46,67	100	
		100	100	100	100	100	100	
		13,68	20	13,68	13,68	11,58	22,11	47,37
1. Púb.	2. Inf. [10,14]	8	8	2	7	6	5	18
		44,44	44,44	11,11	38,89	33,33	27,78	100
		47,06	32	25	41,18	37,5	29,41	36
	1. Sup. [14,20]	8,42	8,42	2,11	7,37	6,32	5,26	18,95
		9	17	6	10	10	12	32
		28,13	53,13	18,75	31,25	31,25	37,50	100
Total	52,94	68	75	58,82	62,5	70,59	64	
	9,47	17,89	6,32	10,53	10,53	12,63	33,68	
	17	25	8	17	16	17	50	
	34	50	16	34	32	34	100	
	100	100	100	100	100	100	100	
	17,89	26,32	8,42	17,89	16,84	17,89	52,63	
Total	2. Inf. [10,14]	17	16	6	15	10	14	39
		43,59	41,03	15,38	38,46	25,64	35,90	100
		56,67	36,36	28,57	50	37,04	36,84	41,05
	1. Sup. [14,20]	17,89	16,84	6,32	15,79	10,53	14,74	41,05
		13	28	15	15	17	24	56
		23,21	50	26,79	26,79	30,36	42,86	100
Total	43,33	63,64	71,43	50	62,96	63,16	58,95	
	13,68	29,47	15,79	15,79	17,89	25,26	58,95	
	30	44	21	30	27	38	95	
	31,58	46,32	22,11	31,58	28,42	40	100	
	100	100	100	100	100	100	100	
	31,58	46,32	22,11	31,58	28,42	40	100	

Nº - % Columna - % Fila - % Total

El cuadro 3 muestra que se cuenta con 112 alumnos de la Licenciatura en Contaduría, 40,18% proveniente de planteles privados y 59,82% de planteles públicos. 41,96% de los alumnos llegan a la universidad con promedios bajos en el bachillerato, de ellos, 12,77% logra promedios elevados en la universidad y

42,55% logra un nivel elevado de eficiencia. Por otra parte, 58,04% de los alumnos llega con promedios altos en el bachillerato, y de ellos, 38,46% logra promedios elevados en la universidad y 60,00% logra un nivel elevado de eficiencia.

Cuadro 3
Contingencia para PromU y Efi: Contaduría

SectorP	PromBR	PromU			Efi			Total
		3. Rep. [0,10)	2. Inf. [10,14)	1. Sup. [14,20]	3. Inf. [0,65)%	2. Med. [65,85)%	1. Sup. [85,100)%	
2. Priv./Otro	2. Inf. [10,14)	5	9	1	5	5	5	15
		33,33	60	6,67	33,33	33,33	33,33	100
		71,43	33,33	9,09	62,5	33,33	22,73	33,33
		4,46	8,04	0,89	4,46	4,46	4,46	13,39
	1. Sup. [14,20]	2	18	10	3	10	17	30
		6,67	60	33,33	10	33,33	56,67	100
		28,57	66,67	90,91	37,5	66,67	77,27	66,67
	Total	1,79	16,07	8,93	2,68	8,93	15,18	26,79
		7	27	11	8	15	22	45
		15,56	60	24,44	17,78	33,33	48,89	100
		100	100	100	100	100	100	
		6,25	24,11	9,82	7,14	13,39	19,64	40,18
1. Púb.	2. Inf. [10,14)	9	18	5	10	7	15	32
		28,13	56,25	15,63	31,25	21,88	46,88	100
		69,23	52,94	25	66,67	46,67	40,54	47,76
		8,04	16,07	4,46	8,93	6,25	13,39	28,57
	1. Sup. [14,20]	4	16	15	5	8	22	35
		11,43	45,71	42,86	14,29	22,86	62,86	100
		30,77	47,06	75	33,33	53,33	59,46	52,24
	Total	3,57	14,29	13,39	4,46	7,14	19,64	31,25
		13	34	20	15	15	37	67
		19,4	50,75	29,85	22,39	22,39	55,22	100
		100	100	100	100	100	100	
		11,61	30,36	17,86	13,39	13,39	33,04	59,82
Total	2. Inf. [10,14)	14	27	6	15	12	20	47
		29,79	57,45	12,77	31,91	25,53	42,55	100
		70	44,26	19,35	65,22	40	33,9	41,96
		12,5	24,11	5,36	13,39	10,71	17,86	41,96
	1. Sup. [14,20]	6	34	25	8	18	39	65
		9,23	52,31	38,46	12,31	27,69	60,00	100
		30	55,74	80,65	34,78	60	66,1	58,04
	Total	5,36	30,36	22,32	7,14	16,07	34,82	58,04
		20	61	31	23	30	59	112
		17,86	54,46	27,68	20,54	26,79	52,68	100
		100	100	100	100	100	100	
		17,86	54,46	27,68	20,54	26,79	52,68	100

N° - % Columna - % Fila - % Total

El cuadro 4 incluye 49 alumnos de Economía, 26,53% de planteles privados y 73,47% de planteles públicos. 46,94% de los alumnos llega con promedios bajos en el bachillerato, de ellos, 17,39% logra promedios

elevados en la universidad y 43,48% un nivel elevado de eficiencia. Por otra parte, 53,06% llega con promedios altos, y de ellos, 34,62% logra promedios elevados en la universidad y 65,38% logra un nivel elevado de eficiencia.

Cuadro 4
Contingencia para PromU y Efi: Economía

SectorP	PromBR	PromU			Efi			Total
		3. Rep. [0,10)	2. Inf. [10,14)	1. Sup. [14,20]	3. Inf. [0,65)%	2. Med. [65,85)%	1. Sup. [85,100)%	
2. Priv./Otro	2. Inf. [10,14)	, 2	2		, ,		4	4
		, 50	50		, ,		100	100
		, 40	28,57		, ,		33,33	30,77
	1. Sup. [14,20]	, 4,08	4,08		, ,		8,16	8,16
		1 3	5		1 ,		8	9
		11,11 33,33	55,56		11,11 ,		88,89	100
	Total	100 60	71,43		100 ,		66,67	69,23
		2,04 6,12	10,2		2,04 ,		16,33	18,37
		1 5	7		1 ,		12	13
1. Púb.	2. Inf. [10,14)	7,69 38,46	53,85		7,69 ,		92,31	100
		100 100	100		100 ,		100	100
		2,04 10,2	14,29		2,04 ,		24,49	26,53
	1. Sup. [14,20]	6 11	2		8 5		6	19
		31,58 57,89	10,53		42,11 26,32		31,58	100
		100 45,83	33,33		72,73 50		40	52,78
	Total	12,24 22,45	4,08		16,33 10,2		12,24	38,78
		, 13	4		3 5		9	17
		, 76,47	23,53		17,65 29,41		52,94	100
Total	, 54,17	66,67		27,27 50		60	47,22	
	, 26,53	8,16		6,12 10,2		18,37	34,69	
	6 24	6		11 10		15	36	
Total	2. Inf. [10,14)	16,67 66,67	16,67		30,56 27,78		41,67	100
		100 100	100		100 100		100	100
		12,24 48,98	12,24		22,45 20,41		30,61	73,47
	1. Sup. [14,20]	6 13	4		8 5		10	23
		26,09 56,52	17,39		34,78 21,74		43,48	100
		85,71 44,83	30,77		66,67 50		37,04	46,94
	Total	12,24 26,53	8,16		16,33 10,2		20,41	46,94
		1 16	9		4 5		17	26
		3,85 61,54	34,62		15,38 19,23		65,38	100
Total	14,29 55,17	69,23		33,33 50		62,96	53,06	
	2,04 32,65	18,37		8,16 10,2		34,69	53,06	
	7 29	13		12 10		27	49	
Total	14,29 59,18	26,53		24,49 20,41		55,1	100	
	100 100	100		100 100		100	100	
	14,29 59,18	26,53		24,49 20,41		55,1	100	

N° - % Columna - % Fila - % Total

El cuadro 5 muestra 31 estudiantes de la licenciatura en estadística, 35,48% de planteles privados y 64,52% de planteles públicos. 48,39% de los alumnos llega con promedios bajos en el bachillerato, de ellos, ninguno (0,00%) logró promedios elevados en la

universidad y 33,33% logró un nivel elevado de eficiencia. Por otra parte, 51,61% llega con promedios altos, de ellos, apenas un 6,25% logra promedios elevados en la universidad y 31,25% logra un nivel elevado de eficiencia.

Cuadro 5
Contingencia para PromU y Efi: Estadística

SectorP	PromBR	PromU			Efi			Total
		3. Rep. [0,10]	2. Inf. [10,14]	1. Sup. [14,20]	3. Inf. [0,65)%	2. Med. [65,85)%	1. Sup. [85,100)%	
2. Priv./Otro	2. Inf. [10,14)	2	2	,	2	1	1	4
		50	50	,	50	25	25	100
		50	33,33	,	28,57	100	33,33	36,36
	1. Sup. [14,20]	6,45	6,45	,	6,45	3,23	3,23	12,9
		2	4	1	5	,	2	7
		28,57	57,14	14,29	71,43	,	28,57	100
	Total	50	66,67	100	71,43	,	66,67	63,64
		6,45	12,9	3,23	16,13	,	6,45	22,58
		4	6	1	7	1	3	11
1. Púb.	2. Inf. [10,14)	36,36	54,55	9,09	63,64	9,09	27,27	100
		100	100	100	100	100	100	100
		12,9	19,35	3,23	22,58	3,23	9,68	35,48
	1. Sup. [14,20]	3	8	,	6	1	4	11
		27,27	72,73	,	54,55	9,09	36,36	100
		60	53,33	,	60	33,33	57,14	55
	Total	9,68	25,81	,	19,35	3,23	12,9	35,48
		2	7	,	4	2	3	9
		22,22	77,78	,	44,44	22,22	33,33	100
Total	40	46,67	,	40	66,67	42,86	45	
	6,45	22,58	,	12,9	6,45	9,68	29,03	
	5	15	,	10	3	7	20	
Total	2. Inf. [10,14)	25	75	,	50	15	35	100
		100	100	,	100	100	100	100
		16,13	48,39	,	32,26	9,68	22,58	64,52
	1. Sup. [14,20]	5	10	,	8	2	5	15
		33,33	66,67	,	53,33	13,33	33,33	100
		55,56	47,62	,	47,06	50	50	48,39
	Total	16,13	32,26	,	25,81	6,45	16,13	48,39
		4	11	1	9	2	5	16
		25	68,75	6,25	56,25	12,5	31,25	100
Total	44,44	52,38	100	52,94	50	50	51,61	
	12,9	35,48	3,23	29,03	6,45	16,13	51,61	
	9	21	1	17	4	10	31	
Total	29,03	67,74	3,23	54,84	12,9	32,26	100	
	100	100	100	100	100	100	100	
	29,03	67,74	3,23	54,84	12,9	32,26	100	

Nº - % Columna - % Fila - % Total

El cuadro 6 señala las cifras para la muestra en su totalidad. Se cuenta con 287 alumnos, 39,72% de planteles privados y 60,28% de planteles públicos. 43,21% llega con promedios bajos en el bachillerato, de ellos, 12,90% logra promedios elevados en la universidad y 39,52% logra un nivel elevado

de eficiencia. Por otra parte, 56,79% llega con promedios altos, y de ellos, un 30,67% logra promedios elevados en la universidad y 52,15% logra un nivel elevado de eficiencia.

Cuadro 6
Contingencia para PromU y Efi: Total

SectorP	PromBR	PromU			Efi			Total
		3. Rep. [0,10)	2. Inf. [10,14)	1. Sup. [14,20]	3. Inf. [0,65)%	Med. [65,85)%	1. Sup. [85,100)%	
2. Priv./Otro	2. Inf. [10,14)	16	21	7	15	10	19	44
		36,36	47,73	15,91	34,09	22,73	43,18	100
		64	36,84	21,88	51,72	37,04	32,76	38,6
	1. Sup. [14,20]	5,57	7,32	2,44	5,23	3,48	6,62	15,33
		9	36	25	14	17	39	70
		12,86	51,43	35,71	20	24,29	55,71	100
	Total	36	63,16	78,13	48,28	62,96	67,24	61,4
		3,14	12,54	8,71	4,88	5,92	13,59	24,39
		25	57	32	29	27	58	114
		21,93	50	28,07	25,44	23,68	50,88	100
1. Púb.	2. Inf. [10,14)	100	100	100	100	100	100	100
		8,71	19,86	11,15	10,1	9,41	20,21	39,72
		26	45	9	31	19	30	80
	1. Sup. [14,20]	32,5	56,25	11,25	38,75	23,75	37,5	100
		63,41	45,92	26,47	58,49	43,18	39,47	46,24
		9,06	15,68	3,14	10,8	6,62	10,45	27,87
	Total	15	53	25	22	25	46	93
		16,13	56,99	26,88	23,66	26,88	49,46	100
		36,59	54,08	73,53	41,51	56,82	60,53	53,76
		5,23	18,47	8,71	7,67	8,71	16,03	32,4
Total	41	98	34	53	44	76	173	
	23,7	56,65	19,65	30,64	25,43	43,93	100	
	100	100	100	100	100	100	100	
	14,29	34,15	11,85	18,47	15,33	26,48	60,28	
	42	66	16	46	29	49	124	
Total	2. Inf. [10,14)	33,87	53,23	12,90	37,1	23,39	39,52	100
		63,64	42,58	24,24	56,1	40,85	36,57	43,21
		14,63	23	5,57	16,03	10,1	17,07	43,21
	1. Sup. [14,20]	24	89	50	36	42	85	163
		14,72	54,6	30,67	22,09	25,77	52,15	100
		36,36	57,42	75,76	43,9	59,15	63,43	56,79
	Total	8,36	31,01	17,42	12,54	14,63	29,62	56,79
		66	155	66	82	71	134	287
		23	54,01	23	28,57	24,74	46,69	100
		100	100	100	100	100	100	100
Total	23	54,01	23	28,57	24,74	46,69	100	
	23	54,01	23	28,57	24,74	46,69	100	

N° - % Columna - % Fila - % Total

4.3. Selección de la muestra

Los 287 estudiantes de los que se compone la muestra fueron seleccionados con base en la información disponible en la base de datos consolidada del Sistema Integrado de Registros Estudiantiles (ULA, SIRE) y el Sistema Central de Registros Estudiantiles (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.

Dichos datos fueron obtenidos gracias a la colaboración de la Dirección de Servicios de Información Administrativa de la ULA, en la persona de su Coordinador de Desarrollo, ingeniero Gustavo Briceño Torres, y constituyen una fuente oficial de la información sobre inscripciones estudiantiles semestre a semestre, así como de los datos de ingreso y egreso de los alumnos a la universidad.

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 a la FACES-ULA se remonta al año 1974 y llega hasta el presente. Para los efectos de este trabajo se considera el año de ingreso como aquel en que el alumno cursó por primera vez una materia en la universidad. Existe información de 10.584 alumnos de la FACES, sin embargo, en conversaciones sostenidas con quienes la proporcionaron, se ha concluido que la máxima confiabilidad se obtendría en aquellos registros de alumnos cuyo ingreso haya ocurrido desde el año 2003 al 2007. Esto reduce los datos para el estudio a un total de 2.252 alumnos. Adicionalmente, no se considerarán las asig-

naturas retiradas legalmente por el alumno, ni aquellas cuya calificación haya sido asentada estando el alumno incurso 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 el sector al que pertenece el instituto de educación media en que egresaron, quedaron 341 estudiantes disponibles.

De ellos, con base en un estudio preliminar de correlación de Pearson (ver gráficos 1) así como a las conclusiones también preliminares obtenidas por Varela, Sinha, Ponsot y Valera (2008), se decidió excluir aquellos alumnos que no hubiesen alcanzado al menos 6 asignaturas cursadas, cantidad que implica en la mayoría de los casos que el alumno ha superado el primer semestre regular de sus estudios.

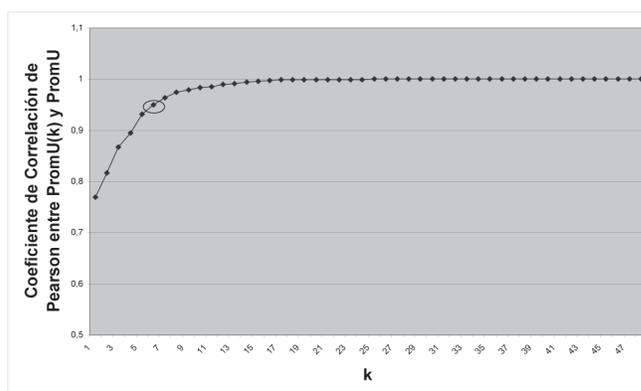


Gráfico 1. Tendencia de correlación en promedios parciales

Este estudio de correlación fue realizado por pares de variables entre $PromU_k$ y $PromU$ (para $k = 1, 2, \dots, N$), donde $PromU_k$ representa el promedio parcial calculado para un alumno, incluyendo las k primeras calificaciones obtenidas en la universidad, hasta incluir el número total de calificaciones con que cuenta (N), variable para cada alumno. Todas las correlaciones entre promedios parciales y promedio final resultan, como cabría esperar,

altamente significativas ($p < 0.001$) frente al análisis de varianza. La correlación es positiva y aumenta a medida que se incluyen más asignaturas en el cálculo del promedio parcial (como muestra la figura). Por último, se puede observar que, a partir de un promedio de calificaciones parcial que incluya más de cinco asignaturas, la correlación supera el valor 0,95, muy cercano al máximo (1,00), por lo cual se supone que en tal caso las variables PromU y Efi resultan comparables entre alumnos.

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 de 287 alumnos con la cual se cuenta para el estudio.

4.4. Cuadro de contingencia para el turno y sector del plantel

El cuadro de contingencia 7 contiene las frecuencias para carreras y sector *versus* turno del plantel. Existen considerablemente más alumnos provenientes de planteles diurnos (D, 86,06%) que de planteles nocturnos u otros (N u O, 13,94%), además, la mayoría de los planteles públicos son de naturaleza diurna. En los privados se encuentra una mayor presencia de planteles nocturnos o de otra naturaleza.

Cuadro 7
Contingencia para el turno del plantel

Carrera	SectorP	Turno del Plantel		
		D	N u O	Total
Administración	2. Priv./Otro	34	11	45
		75,56	24,44	100
		41,98	78,57	47,37
		11,85	3,83	15,68
	1. Púb.	47	3	50
		94	6	100
		58,02	21,43	52,63
		16,38	1,05	17,42
	Total	81	14	95
		85,26	14,74	100
Contaduría	2. Priv./Otro	33	12	45
		73,33	26,67	100
		35,11	66,67	40,18
		11,5	4,18	15,68
	1. Púb.	61	6	67
		91,04	8,96	100
		64,89	33,33	59,82
		21,25	2,09	23,34
	Total	94	18	112
		83,93	16,07	100
Economía	2. Priv./Otro	11	2	13
		84,62	15,38	100
		23,91	66,67	26,53
		3,83	0,7	4,53
	1. Púb.	35	1	36
		97,22	2,78	100
		76,09	33,33	73,47
		12,2	0,35	12,54
	Total	46	3	49
		93,88	6,12	100
Total	100	100	100	
	16,03	1,05	17,07	

Carrera	SectorP	Turno del Plantel		
		D	N u O	Total
Estadística	2. Priv./Otro	6	5	11
		54,55	45,45	100
		23,08	100	35,48
		2,09	1,74	3,83
	1. Púb.	20	,	20
		100	,	100
		76,92	,	64,52
		6,97	,	6,97
	Total	26	5	31
		83,87	16,13	100
Total	2. Priv./Otro	84	30	114
		73,68	26,32	100
		34,01	75	39,72
		29,27	10,45	39,72
	1. Púb.	163	10	173
		94,22	5,78	100
		65,99	25	60,28
		56,79	3,48	60,28
	Total	247	40	287
		86,06	13,94	100
Total	100	100	100	
	86,06	13,94	100	

Nº - % Columna - % Fila - % Total

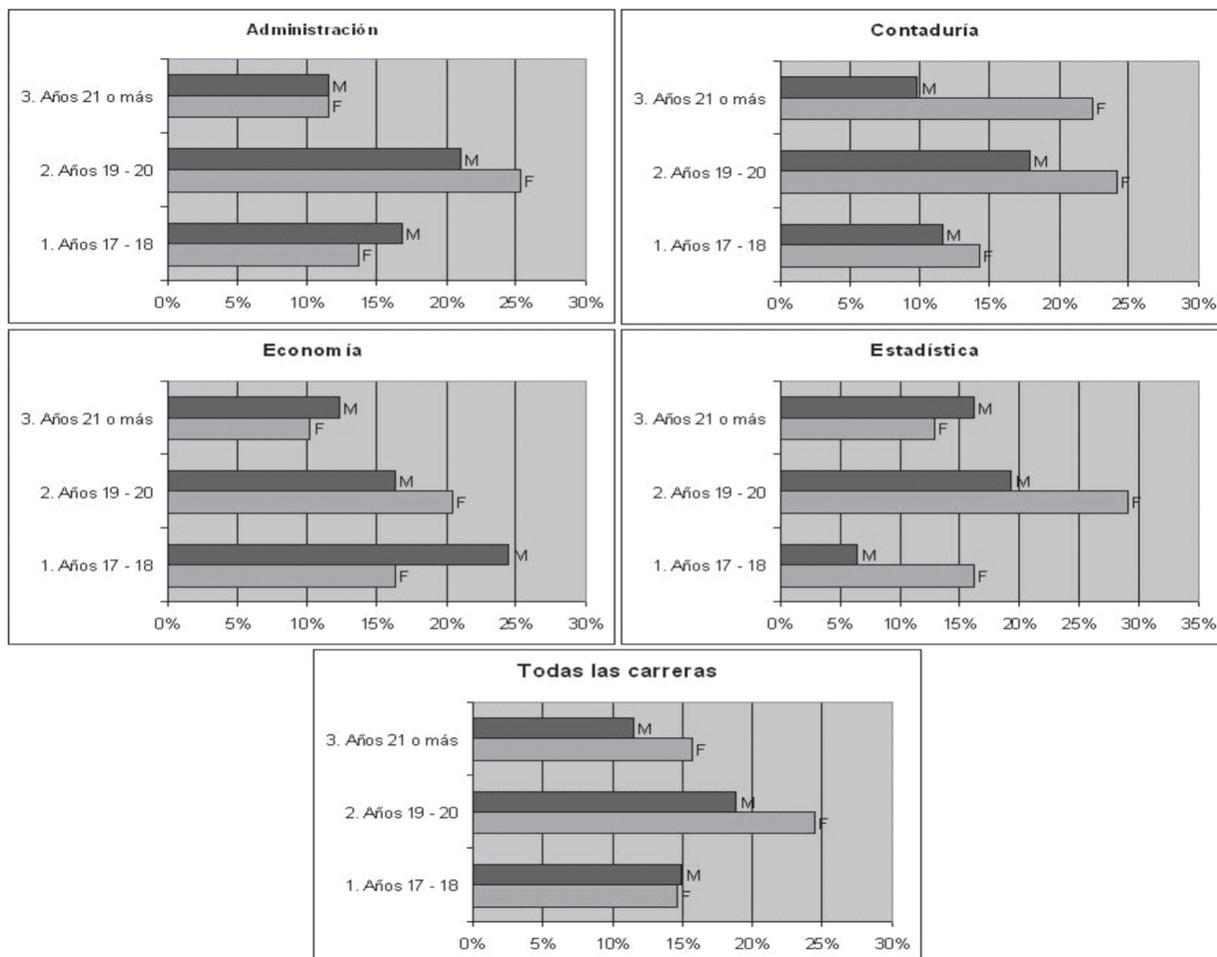


Gráfico 2. Distribución por edad y sexo en la muestra

4.5. Distribución por edad y sexo

El gráfico 2 muestra la distribución por edad y sexo para las carreras y en general.

Como puede apreciarse, los datos reproducen la composición esperada, esto es, más alumnas que alumnos (excepto en Economía) y mayoritariamente con edades inferiores a los 20 años al ingreso.

4.6. Comportamiento de los promedios y la eficiencia

El gráfico 2 muestra el diagrama de caja para las dos variables de promedio (PromU

y PromBR) y la variable de eficiencia (Efi). Atendiendo al objetivo 1, la primera y la tercera se consideran variables respuesta y la segunda es una variable explicativa y, atendiendo al objetivo 2, esta última es también una variable respuesta. Puede notarse que los promedios de calificaciones de bachillerato son semejantes en la muestra, un tanto superiores a los 14 puntos. El rango más amplio para esta variable aparece en la carrera de Administración de Empresas. Por otra parte, los promedios obtenidos en la universidad se reducen ostensiblemente

(ahora por debajo de los 13 puntos) y su rango incluye el límite inferior (0 puntos) mas no el superior (20 puntos). De las cuatro carreras en consideración, Estadística es la que cuenta

con los promedios de calificaciones más bajos, y es bastante similar el comportamiento de las restantes.

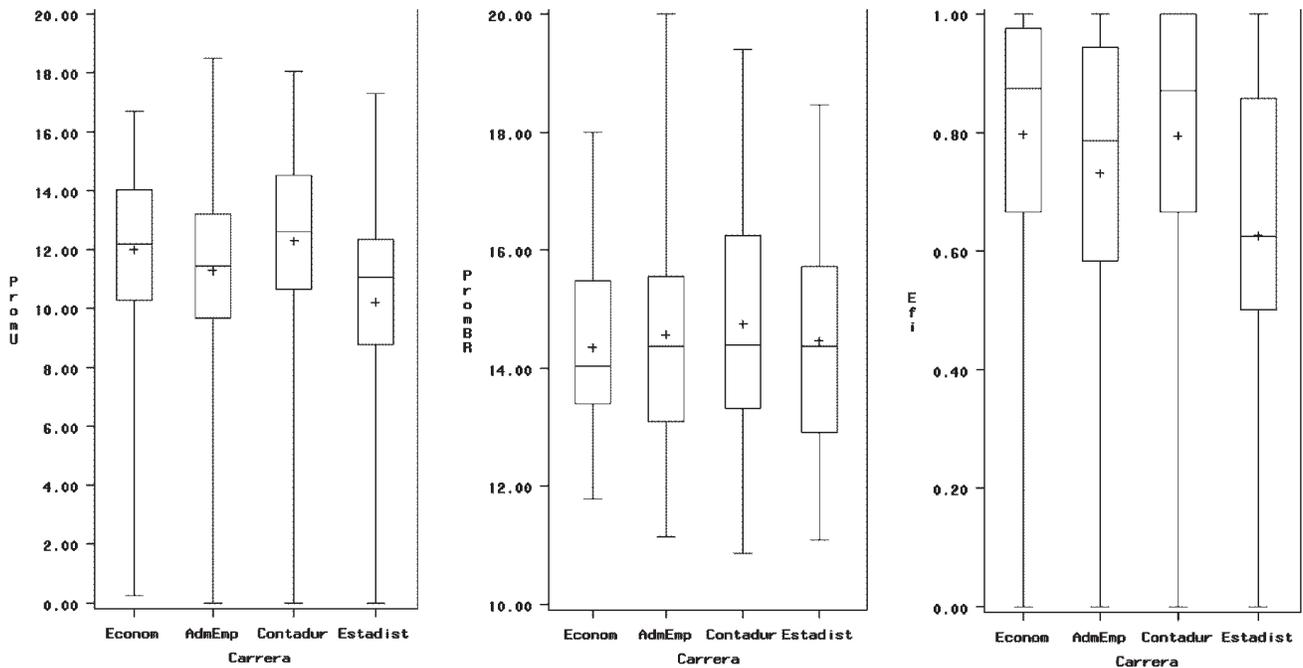


Gráfico 3. Boxplot para PromBR, PromU y Efi

Sobre la eficiencia, puede decirse que su comportamiento es equivalente al del promedio universitario y la mayoría de los valores caen por encima del 50%, es decir que, en promedio, de cada dos asignaturas que inscriben y cursan los alumnos de la muestra, aprueban cuando menos una de ellas.

5. Modelo de regresión logística

Esta sección presenta las ecuaciones del modelo de regresión logística para la variable respuesta PromU y las variables explicativas

SectorP y PromBR. La versión final empleada en el trabajo, al contener un mayor número de variables explicativas, resulta una extensión completamente análoga.

El modelo estadístico escogido para el análisis de los datos es el *modelo de regresión logística con respuesta policotómica ordinal* (Stokes *et al.*, 2000:243), también conocido como *modelo de posibilidades proporcionales* o simplemente *logit acumulado*, originalmente propuesto por McCullagh (1980).

Este modelo del tipo lineal generalizado (Rodríguez, 2008) supone una distribución

multinomial en los datos y una función de enlace *logit*, canónica, esto es, $\log(\pi / (1-\pi))$, también conocida como logaritmo de posibilidad, ya que si π es la probabilidad de éxito de un evento, la cantidad $\pi / (1-\pi)$ representa el número de veces en que es más (o menos) posible que ocurra un éxito frente a un fracaso. Véanse los tres ejemplos que siguen: (1) Para $\pi = 0,5$, $\pi / (1-\pi) = 1$, la interpretación es que un éxito tiene la misma posibilidad de ocurrir que un fracaso; (2) para $\pi = 0,8$, $\pi / (1-\pi) = 4$, la interpretación es que un éxito tiene cuatro veces más posibilidades de ocurrir que un fracaso, y (3) para $\pi = 0,3$, $\pi / (1-\pi) = 0,43$ y $[\pi / (1-\pi)]^{-1} = 2,33$, la interpretación es que un fracaso tiene 2,33 veces más posibilidades de ocurrir que un éxito. Una discusión amplia

de esta familia de modelos puede verse en Agresti (2007).

El modelo de posibilidades proporcionales se utiliza cuando la variable respuesta tiene más de dos niveles o categorías, pero dichos niveles pueden ser entendidos siguiendo un cierto orden de precedencia. Tal es el caso con las variables respuesta de interés primordial (PromU y Efi), como puede apreciarse en la tabla 1.

El cuadro 8 presenta la situación de las variables de interés, donde $n_{i\bullet} = \sum_{k=1}^3 n_{ik}$, para $i = 1, 2, 3, 4$. Nótese que se ha preferido emplear un sólo índice en lugar de dos para representar las combinaciones de niveles de las variables explicativas.

Cuadro 8
Variable de interés para el modelo ilustrativo

PromBR	SectorP	PromU			Total
		3. Rep. [0,10)	2. Inf. [10,14)	1. Sup. [14,20]	
2. Inf. [10,14)	2. Priv./Otro	n_{43}	n_{42}	n_{41}	$n_{4\bullet}$
	1. Púb.	n_{33}	n_{32}	n_{31}	$n_{3\bullet}$
1. Sup. [14,20]	2. Priv./Otro	n_{23}	n_{22}	n_{21}	$n_{2\bullet}$
	1. Púb.	n_{13}	n_{12}	n_{11}	$n_{1\bullet}$

El nivel de referencia contra el cual se realiza el análisis de posibilidades es el primero que aparece en el cuadro de contingencia, es decir, el más bajo o inferior en la escala (2 para PromBR y SectorP, y 3 para PromU). Sea π_{i1} la probabilidad de un rendimiento superior (en el intervalo [14, 20] puntos en promedio), π_{i2} la probabilidad de un rendimiento inferior (en el intervalo [10, 14) puntos en promedio) y π_{i3} la probabilidad de un rendimiento reprobatorio (en el intervalo [0, 10) puntos en promedio), para la i -ésima combinación de niveles de PromBR y SectorP ($i = 1, 2, 3, 4$). Por ejemplo,

$i = 1$ hace referencia a PromBR "1. Sup. [14,20]" y SectorP "1. Púb.", mientras que $i = 4$ hace referencia a PromBR "2. Inf. [10,14)" y SectorP "2. Priv./Otro". Ahora, sean $\theta_{i1} = \pi_{i1}$ y $\theta_{i2} = \pi_{i1} + \pi_{i2}$. Se requiere que $\sum_{k=1}^3 \pi_{ik} = 1, \forall i$. Son en total $4 \times 2 = 8$ combinaciones no dependientes para θ_{ik} que representan las probabilidades acumuladas de obtener un rendimiento igual o superior a 14 puntos ($k = 1$) y un rendimiento igual o superior a 10 puntos ($k = 2$). Por el complemento es claro que $\pi_{i3} = 1 - \theta_{i2}$, que se asocia con un rendimiento

inferior a los 10 puntos. Así, son necesarias dos funciones *logit* como las que se muestran en las ecuaciones 1 y 2.

$$\text{logit}(\theta_{i1}) = \log \left[\frac{\pi_{i1}}{\pi_{i2} + \pi_{i3}} \right] \quad (1)$$

$$\text{logit}(\theta_{i2}) = \log \left[\frac{\pi_{i1} + \pi_{i2}}{\pi_{i3}} \right] \quad (2)$$

En efecto, sea $\theta_k = \sum_{r=1}^k \pi_r$, para $k=1,2$, entonces:

$$1. \text{logit}(\theta_{i1}) = \log \left[\frac{\sum_{r=1}^1 \pi_r}{1 - \sum_{r=1}^1 \pi_r} \right] = \log \left[\frac{\pi_{i1}}{1 - \pi_{i1}} \right] = \log \left[\frac{\pi_{i1}}{\pi_{i2} + \pi_{i3}} \right]$$

$$2. \text{logit}(\theta_{i2}) = \log \left[\frac{\sum_{r=1}^2 \pi_r}{1 - \sum_{r=1}^2 \pi_r} \right] = \log \left[\frac{\pi_{i1} + \pi_{i2}}{1 - \pi_{i1} - \pi_{i2}} \right] = \log \left[\frac{\pi_{i1} + \pi_{i2}}{\pi_{i3}} \right]$$

Por otra parte, suponiendo posibilidades proporcionales, es decir, suponiendo que los parámetros, a excepción del intercepto, son los mismos para todos los *logits*, el modelo de regresión multinomial emplea también la función de enlace *logit* y se puede escribir de forma equivalente como $\text{logit}(\theta_k) = \alpha_k + x_i' \beta$, o bien como la ecuación 3.

$$\theta_k = \frac{\exp\{\alpha_k + \sum_{r=1}^2 x_r \beta_r\}}{1 + \exp\{\alpha_k + \sum_{r=1}^2 x_r \beta_r\}}, \quad \text{para } k=1,2 \quad (3)$$

En la ecuación 3, α_1 y α_2 son los interceptos correspondientes a cada *logit*, x_i' es la fila con dos componentes de la submatriz de diseño que corresponde a los β 's, y $\beta' = [\beta_1 \ \beta_2]'$, son los r parámetros necesarios debido a los niveles incrementales de las covariables ($r=2$), esto es: β_1 corresponde al nivel "1. Sup. [14,20]" de PromBR y β_2 corresponde al nivel "1. Púb." de SectorP. En efecto,

$$\text{logit}(\theta_k) = \log \left[\frac{\sum_{r=1}^k \pi_r}{1 - \sum_{r=1}^k \pi_r} \right] = \alpha_k + x_i' \beta \Rightarrow \frac{\sum_{r=1}^k \pi_r}{1 - \sum_{r=1}^k \pi_r} = e^{\{\alpha_k + x_i' \beta\}}$$

, luego

$$\frac{e^{\{\alpha_k + x_i' \beta\}}}{1 + e^{\{\alpha_k + x_i' \beta\}}} = \frac{\left[\frac{\sum_{r=1}^k \pi_r}{1 - \sum_{r=1}^k \pi_r} \right]}{1 + \left[\frac{\sum_{r=1}^k \pi_r}{1 - \sum_{r=1}^k \pi_r} \right]} = \frac{\left[\frac{\sum_{r=1}^k \pi_r}{1 - \sum_{r=1}^k \pi_r} \right]}{\left[\frac{1 - \sum_{r=1}^k \pi_r + \sum_{r=1}^k \pi_r}{1 - \sum_{r=1}^k \pi_r} \right]} = \frac{\left(\sum_{r=1}^k \pi_r \right) \left(1 - \sum_{r=1}^k \pi_r \right)}{\left(1 - \sum_{r=1}^k \pi_r \right) \times 1} = \sum_{r=1}^k \pi_r = \theta_k$$

En notación matricial, el modelo se expresa según la ecuación 4.

$$\begin{bmatrix} \text{logit}(\theta_1) \\ \text{logit}(\theta_2) \\ \text{logit}(\theta_2) \\ \text{logit}(\theta_3) \\ \text{logit}(\theta_3) \\ \text{logit}(\theta_4) \\ \text{logit}(\theta_4) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{bmatrix} \quad (4)$$

Por ejemplo, $\text{logit}(\theta_3) = \alpha_2 + \beta_2$ modela el supuesto de que el intercepto del segundo *logit* aporta a la explicación de la variabilidad en los datos, no así el nivel "1. Sup. [14,20]" de PromBR, pero si lo hace el nivel "1. Púb." de SectorP.

6. Resultados de las regresiones logísticas

6.1. Regresión para PromU

Se propone un modelo de regresión logística de posibilidades proporcionales en el cual la variable respuesta es PromU y las variables explicativas son Carrera, SectorP, PromBR, Edad, Sexo y Turno, como se describen en el cuadro 1. Este modelo, análogo al expuesto en la sección anterior considera sólo los efectos principales y no las interacciones, como puede apreciarse en la ecuación 5.

$$\text{logit}(\theta_k) = \alpha_k + x_i'[\beta_1 \beta_2 \beta_3 \beta_4 \beta_5 \beta_6 \beta_7 \beta_8 \beta_9]$$

$$i = 1, 2, \dots, 192; k = 1, 2$$

Donde i es el índice que representa las combinaciones de niveles de las variables explicativas ($4 \times 2 \times 2 \times 3 \times 2 \times 2 = 192$), k es índice que representa cada una de las dos funciones *logit* necesarias para modelar la variable respuesta PromU, α_k es el k -ésimo intercepto, los parámetros de los efectos correspondientes a los niveles “Administración”, “Contaduría” y “Economía” de la variable Carrera son $\beta_1, \beta_2, \beta_3$ respectivamente, β_4 es el parámetro del nivel “1. Púb” de SectorP, β_5 es el parámetro del nivel “1. Sup. [14,20]” de PromBR, los parámetros de los efectos correspondientes a los niveles “1. Años 17-18”, “2. Años 19-20” de la variable Edad son β_6, β_7 respectivamente, β_8 es el parámetro del nivel “F” de la variable Sexo y β_9 es el parámetro del nivel “D” de la variable Turno.

Con un valor 9,97 de una χ^2 (Chi²) con 9 grados de libertad (GL), no se encuentra evidencia para rechazar la hipótesis de posibilidades proporcionales en el modelo ($p > 0,35$). Los estadísticos de *Deviance* y Pearson, que miden la bondad del ajuste del modelo propuesto, resultan estimados en 217 y 212 respectivamente, con 199 GL, y se encuentran en la región de aceptación de la hipótesis de ajuste ($p > 0,18$ en el primero y $p > 0,25$ en el segundo). No hay problemas de sobre ni sub-dispersión por lo cual no es necesario aplicar corrección alguna (la división entre el valor estimado de *Deviance* y los GL es 1,090, entre el estimado de Pearson y los GL es 1,065). En síntesis, el modelo se ajusta bien a los datos.

El cuadro 9 contiene el análisis de varianza para docimar la hipótesis de nulidad de los efectos presentes en el modelo.

Cuadro 9
ANOVA Tipo III para los efectos principales: PromU

Efecto	GL	Chi ² de Wald	Pr > Chi ²
Carrera	3	9,6291	0,0220
PromBR	1	16,4470	<,0001
SectorP	1	0,7337	0,3917
Edad	2	2,1887	0,3347
sexo	1	2,4251	0,1194
Turno	1	0,0038	0,9505

Con un nivel inferior al 3%, sólo los efectos Carrera y PromBR resultan significativos. Esto implica que el resto de los factores no contribuyen a explicar el rendimiento universitario.

El cuadro 10 contiene las estimaciones de los parámetros en general y posibilidades sólo para los efectos que resultaron significativos (el intercepto no se toma en cuenta para el

análisis). Allí se reproducen los estimadores de los parámetros del modelo propuesto, sin embargo, lo interesante de este análisis está en el estimador de posibilidad. El estimador de posibilidades siempre se confronta contra el último nivel (en este caso, “Estadística” para la variable Carrera y “2. Inf. [10,14]” para PromBR) y se consigue exponenciando el estimador del parámetro.

Cuadro 10
Estimación máximo verosímil de parámetros y posibilidades: PromU

Parámetro	GL	Esti- mador	Error Estándar	Chi ² de Wald		Estimador Posibilidad	Límites del 95% de confianza		
				Pr > Chi ²					
Intercepto	1. Sup. [14,20]	1	-2,4776	0,5492	20,3492	<,0001			
Intercepto	2. Inf. [10,14]	1	0,2159	0,5246	0,1694	0,6806			
Carrera	Administración	1	0,3023	0,4068	0,5523	0,4574	1,3530	0,6100	3,0030
Carrera	Contaduría	1	0,8868	0,4002	4,9095	0,0267	2,4270	1,1080	5,3190
Carrera	Economía	1	1,0467	0,4567	5,2514	0,0219	2,8480	1,1640	6,9720
PromBR	1. Sup. [14,20]	1	1,0493	0,2587	16,4470	<,0001	2,8560	1,7200	4,7410
SectorP	1. Púb.	1	-0,2142	0,2500	0,7337	0,3917			
Edad	1. Años 17-18	1	-0,0062	0,3235	0,0004	0,9846			
Edad	2. Años 19-20	1	-0,3496	0,2807	1,5508	0,2130			
Sexo	F	1	0,3765	0,2418	2,4251	0,1194			
Turno	D	1	-0,0217	0,3492	0,0038	0,9505			

Las estimaciones de posibilidades son las siguientes:

1. Un alumno de la Licenciatura en Contaduría tiene 2,43 veces más posibilidades de obtener un rendimiento "1. Sup. [14,20]" versus {"2. Inf. [10,14]" o "3. Rep. [0,10]"}, de las que tiene un alumno de la Licenciatura en Estadística. También tiene 2,43 veces más posibilidades de obtener un rendimiento {"1. Sup. [14,20]" o "2. Inf. [10,14]"} versus "3. Rep. [0,10]", de las que tiene un alumno de la Licenciatura en Estadística.
2. Un alumno de Economía tiene 2,85 veces más posibilidades de obtener un rendimiento "1. Sup. [14,20]" versus {"2. Inf. [10,14]" o "3. Rep. [0,10]"}, de las que tiene un alumno de la Licenciatura en Estadística. Tiene también 2,85 veces más posibilidades de obtener un rendimiento {"1. Sup. [14,20]" o "2. Inf. [10,14]"} versus "3. Rep. [0,10]", de las que tiene un alumno de la Licenciatura en Estadística.
3. Sobre la Licenciatura en Administración no es conveniente realizar estimaciones pues el parámetro resulta no ser significativo.
4. Aun cuando las estimaciones de posibilidad para la comparación entre la Licenciatura en

Contaduría y Economía no están calculadas en la tabla, son sencillas de obtener a partir del modelo. Basta hacer la división entre las tasas de posibilidad de ambos estimadores, esto es, un alumno de Economía tiene $2,848/2,427=1,17$ veces más posibilidades de obtener un rendimiento "1. Sup. [14,20]" versus {"2. Inf. [10,14]" o "3. Rep. [0,10]"}, de las que tiene un alumno de la Licenciatura en Contaduría. Tiene también 1,17 veces más posibilidades de obtener un rendimiento {"1. Sup. [14,20]" o "2. Inf. [10,14]"} versus "3. Rep. [0,10]", de las que tiene un alumno de la Licenciatura en Contaduría.

5. Finalmente, lo medular, un alumno cuyo promedio en la secundaria sea "1. Sup. [14,20]" tiene, con certeza estadística, 2,86 veces más posibilidades de obtener un rendimiento "1. Sup. [14,20]" versus {"2. Inf. [10,14]" o "3. Rep. [0,10]"}, de las que tiene un alumno cuyo promedio de calificaciones en la secundaria sea "2. Inf. [10,14]". También tiene 2,86 veces más posibilidades de obtener un rendimiento {"1. Sup. [14,20]" o "2. Inf. [10,14]"} versus "3. Rep. [0,10]", de las que tiene un alumno cuyo promedio de calificaciones en la secundaria sea "2. Inf. [10,14]"

6.2. Regresión para Efi

Se propone un modelo de regresión logística de posibilidades proporcionales en el cual la variable respuesta es Efi y las variables explicativas son Carrera, SectorP, PromBR, Edad, Sexo y Turno, como antes. Este modelo es muy similar al expuesto en la sección anterior, excepto por el cambio en la variable respuesta. Como ambas variables respuesta tienen el mismo número de niveles (3 en total), el modelo corresponde exactamente a la ecuación 5.

Con un valor 8,31 de una χ^2 con 9 GL, no se encuentra evidencia para rechazar la hipótesis de posibilidades proporcionales en el modelo ($p > 0,50$). Así también, los estadísticos de *Deviance* y Pearson se estiman en 219 y 191 respectivamente, con 199 GL, y se encuentran en la región de aceptación de

la hipótesis de ajuste ($p > 0,16$ en el primero y $p > 0,65$ en el segundo). En este caso, hay un leve problema de sobre-dispersión en el estadístico *Deviance* y de sub-dispersión en el estadístico de Pearson (la división entre el valor estimado de *Deviance* y los GL es 1,1014, entre el estimado de Pearson y los GL es 0,9590). Se decide entonces escalar *Deviance*. Este modelo también se ajusta bien a los datos.

El cuadro 11 contiene el análisis de varianza para docimar la hipótesis de nulidad de los efectos presentes en el modelo.

Con un nivel inferior al 2%, nuevamente sólo los efectos Carrera y PromBR resultan significativos. El resto de los factores no contribuyen a explicar la eficiencia en los estudios universitarios.

Cuadro 11
ANOVA Tipo III para los efectos principales: Efi

Efecto	GL	Chi ² de Wald	Pr > Chi ²
Carrera	3	12,0366	0,0073
PromBR	1	5,6773	0,0172
SectorP	1	1,3340	0,2481
Edad	2	1,1088	0,5744
sexo	1	2,0083	0,1564
Turno	1	0,2744	0,6004

Cuadro 12
Estimación máximo verosímil de parámetros y posibilidades: Efi

Parámetro		GL	Estimador	Error Estándar	Chi ² de Wald	Pr > Chi ²	Estimador Posibilidad	Límites del 95% de confianza	
Intercepto	1. Sup. [85,100]%	1	-1,7579	0,5485	10,2698	0,0014			
Intercepto	2. Med. [65,85]%	1	-0,6283	0,5384	1,3617	0,2432			
Carrera	Administración	1	0,7020	0,4148	2,8641	0,0906	2,0180	0,8950	4,5490
Carrera	Contaduría	1	1,2107	0,4103	8,7085	0,0032	3,3560	1,5020	7,4990
Carrera	Economía	1	1,3677	0,4709	8,4368	0,0037	3,9260	1,5600	9,8800
PromBR	1. Sup. [14,20]	1	0,6125	0,2571	5,6773	0,0172	1,8450	1,1150	3,0540
SectorP	1. Púb.	1	-0,2991	0,2590	1,3340	0,2481			
Edad	1. Años 17-18	1	0,3416	0,3353	1,0378	0,3083			
Edad	2. Años 19-20	1	0,0966	0,2875	0,1128	0,7370			
Sexo	F	1	0,3493	0,2465	2,0083	0,1564			
Turno	D	1	0,1871	0,3572	0,2744	0,6004			

El cuadro 12 contiene las estimaciones de los parámetros en general y posibilidades sólo para los efectos que resultaron significativos. Como antes, el estimador de posibilidades siempre se confronta contra el último nivel.

Las estimaciones de posibilidades son las siguientes:

1. En este caso, el nivel Administración tampoco resulta significativo. No obstante, como su significación es inferior al 10%, se le incluirá en el análisis, advirtiendo que los resultados pueden no ser confiables. Así, un alumno de la Licenciatura en Administración tiene 2,02 veces más posibilidades de obtener una eficiencia "1. Sup. [85,100]%" versus {"2. Med. [65,85]%" o "3. Inf. [0,65]%"}, de las que tiene un alumno de la Licenciatura en Estadística. Tiene también 2,02 veces más posibilidades de obtener un rendimiento {"1. Sup. [85,100]%" o "2. Med. [65,85]%" } versus "3. Inf. [0,65]%", de las que tiene un alumno de la Licenciatura en Estadística.
2. Un alumno de la Licenciatura en Contaduría tiene 3,36 veces más posibilidades de obtener una eficiencia "1. Sup. [85,100]%" versus {"2. Med. [65,85]%" o "3. Inf. [0,65]%"}, de las

que tiene un alumno de la Licenciatura en Estadística. Tiene también 3,36 veces más posibilidades de obtener un rendimiento {"1. Sup. [85,100]%" o "2. Med. [65,85]%" versus "3. Inf. [0,65]%"}, de las que tiene un alumno de la Licenciatura en Estadística.

3. Un alumno de Economía tiene 3,93 veces más posibilidades de obtener una eficiencia "1. Sup. [85,100]%" versus {"2. Med. [65,85]%" o "3. Inf. [0,65]%"}, de las que tiene un alumno de la Licenciatura en Estadística. Tiene también 3,93 veces más posibilidades de obtener un rendimiento {"1. Sup. [85,100]%" o "2. Med. [65,85]%" } versus "3. Inf. [0,65]%", de las que tiene un alumno de la Licenciatura en Estadística.
4. Un alumno de Economía tiene $3,93/2,02=1,95$ veces más posibilidades de obtener una eficiencia "1. Sup. [85,100]%" versus {"2. Med. [65,85]%" o "3. Inf. [0,65]%"}, de las que tiene un alumno de la Licenciatura en Administración. Tiene también 1,95 veces más posibilidades de obtener un rendimiento {"1. Sup. [85,100]%" o "2. Med. [65,85]%" } versus "3. Inf. [0,65]%", de las que tiene un alumno de la Licenciatura en Administración.

5. Un alumno de Economía tiene $3,93/3,36=1,17$ veces más posibilidades de obtener una eficiencia "1. Sup. [85,100]%" versus {"2. Med. [65,85]%" o "3. Inf. [0,65]%"}, de las que tiene un alumno de la Licenciatura en Contaduría. Tiene también 1,17 veces más posibilidades de obtener un rendimiento {"1. Sup. [85,100]%" o "2. Med. [65,85]%" } versus "3. Inf. [0,65]%", de las que tiene un alumno de la Licenciatura en Contaduría.
6. Un alumno de la Licenciatura en Contaduría tiene $3,36/2,02=1,66$ veces más posibilidades de obtener una eficiencia "1. Sup. [85,100]%" versus {"2. Med. [65,85]%" o "3. Inf. [0,65]%"}, de las que tiene un alumno de la Licenciatura en Administración. Tiene también 1,66 veces más posibilidades de obtener un rendimiento {"1. Sup. [85,100]%" o "2. Med. [65,85]%" } versus "3. Inf. [0,65]%", de las que tiene un alumno de la Licenciatura en Administración.
7. Finalmente, un alumno cuyo promedio en la secundaria sea "1. Sup. [14,20]" tiene, con certeza estadística, 1,85 veces más posibilidades de obtener una eficiencia "1. Sup. [85,100]%" versus {"2. Med. [65,85]%" o "3. Inf. [0,65]%"}, de las que tiene un alumno cuyo promedio de calificaciones en la secundaria sea "2. Inf. [10,14)". Tiene también 1,85 veces más posibilidades de obtener un rendimiento {"1. Sup. [85,100]%" o "2. Med. [65,85]%" } versus "3. Inf. [0,65]%", de las que tiene un alumno cuyo promedio de calificaciones en la secundaria sea "2. Inf. [10,14)".

6.3. Regresión para PromBR

Se propone un modelo de regresión logística binaria en el cual la variable respuesta es PromBR y las variables explicativas son SectorP y Turno, como se describen en el cuadro 1. Este modelo, análogo a los anteriores, pero mucho más simple, pues la variable respuesta sólo tiene dos niveles, considera los efectos principales y se describe como en la ecuación 6.

$$\text{logit}(\theta_i) = \alpha + x_i'[\beta_1 \ \beta_2] \quad i=1,2,3,4 \quad (6)$$

Donde i es el índice que representa las combinaciones de niveles de las variables explicativas ($2 \times 2 = 4$), es necesaria una sola función *logit* para modelar la variable respuesta PromBR, β_1 es el parámetro del efecto correspondiente al nivel "1. Púb." de SectorP y β_2 es el parámetro del nivel "D" de la variable Turno.

Los estadísticos de *Deviance* y Pearson resultan estimados en 0,7880 y 0,7839 respectivamente, con 1 GL, y se encuentran en la región de aceptación de la hipótesis de buen ajuste del modelo ($p > 0,37$ en el primero y $p > 0,38$ en el segundo), sin embargo, hay problemas de sub-dispersión en ambos y se decide nuevamente escalar *Deviance*. Aún así, el modelo también se ajusta bien a los datos.

El cuadro 13 contiene el análisis de varianza para docimar la hipótesis de nulidad de los efectos presentes en el modelo.

Cuadro 13
ANOVA Tipo III para los efectos principales: PromBR

Efecto	GL	Chi ² de Wald	Pr > Chi ²
SectorP	1	2,1626	0,1414
Turno	1	0,1062	0,7445

Ninguno de los efectos considerados resulta significativo. Como el modelo ajusta correctamente los datos, esto implica que la variabilidad se explica sólo por la acción del intercepto y el promedio de secundaria no tiene relación causal ni con el sector del plantel de procedencia ni con el turno. Tanto la estimación de parámetros como el estudio de posibilidades no resultan de interés en este caso.

7. Conclusiones

Los estudios sobre el desempeño estudiantil en la universidad son siempre una fuente de preocupación social. En Venezuela, lamentablemente, la discusión sobre el tema pocas veces se enfrenta con argumentos científicos que la sustenten y, las más de las veces, los argumentos revisten carácter subjetivo. Este no es el caso del presente trabajo. Los datos han sido obtenidos sin sesgo de ninguna especie y de fuentes oficiales, se han descrito abundantemente y se han probado modelos con valor inferencial.

En primer término, se concluye que los modelos de regresión logística demuestran ser una herramienta muy poderosa para indagar sobre la explicación de variables categóricas de respuesta, en las que el investigador tenga especial interés.

En este caso se puede afirmar, con certidumbre estadística, lo siguiente:

1. Sobre la hipótesis (a): El rendimiento y la eficiencia en una carrera universitaria de la FACES-ULA se explican sólo en función de la carrera que se cursa y del promedio de calificaciones que el alumno obtuvo en sus estudios secundarios. Para el resto de los factores incluidos en la hipótesis no hay evidencia suficiente que permita suponer que influyen en su explicación. En la FACES-ULA los alumnos que obtienen más bajos promedios de calificaciones son los que cursan la Licenciatura

en Estadística, cuando se les compara con alumnos de las restantes carreras. Las posibilidades de que un estudiante con notas promedio de secundaria superiores o iguales a los 14 puntos, obtenga un promedio de notas universitario de aceptable a elevado (también superior a 14 puntos), son casi tres veces más que las que tiene un alumno con bajas notas promedio de secundaria. Un buen alumno de secundaria también tiene casi dos veces más posibilidades de obtener una eficiencia elevada en sus estudios universitarios, de las que tiene un alumno de bajas calificaciones en secundaria.

2. Sobre la hipótesis (b): No hay evidencia para afirmar que alumnos provenientes de planteles públicos y privados, o de turnos distintos, obtengan calificaciones diferentes en sus estudios secundarios.

En cuanto a lo discutido en las secciones iniciales del trabajo, al menos con base en los datos de los alumnos de la FACES-ULA, desde 2003 hasta el 2007, estas conclusiones suponen:

- I. No hay razón académica sostenible para discriminar el ingreso a la universidad, favoreciendo a aquellos alumnos que provengan de planteles del sector público, pues el sector del plantel de procedencia no contribuye en nada a la explicación del rendimiento universitario y
- II. Tampoco hay razón académica sostenible para privilegiar pruebas internas de admisión, que no toman en consideración el promedio de calificaciones obtenido por el alumno en sus estudios secundarios, ya que ha sido demostrado que esta variable si explica, estadísticamente hablando, el rendimiento en la universidad.

Finalmente es menester acotar que, en el futuro, el equipo de trabajo piensa emprender análisis similares que profundicen en dos direcciones importantes: Ampliando las facultades y carreras de análisis, e incluyendo datos sobre resultados de pruebas internas

de admisión. Estas direcciones seguramente complementarán las conclusiones logradas con el presente trabajo e incrementarán la base científica que se acumula sobre el rendimiento estudiantil universitario, que tanta falta hace al país.

8. Referencias

- Agresti, Alan (2007). *"An introduction to categorical data analysis"*. 2° Edition. New Jersey: JohnWiley & Sons, Inc.
- Di Gresia, Luciano; Porto, Alberto y Ripani, Laura (2002). *"Rendimiento de los estudiantes de las universidades públicas argentinas"*. Documento de Trabajo Nro. 45. Consultado el 04-Marzo-2008 en <http://www.depeco.econo.unlp.edu.ar/doctrab/doc45.pdf>
- Fuenmayor Toro, et al. (2002). *"La admisión estudiantil a las universidades públicas venezolanas: causas de las iniquidades"*. Revista de Pedagogía (Caracas, UCV), Vol XXIII-N°66.
- Garnica, Elsy (1997). *"El rendimiento estudiantil: Una metodología para su medición"*. Economía (Mérida, Venezuela), N° 13.
- Garnica, Elsy; González, Pilar; Díaz, Amelia y Torres, Enrique (1991). *"Análisis discriminante. Estudio del rendimiento estudiantil."* Economía (Mérida, Venezuela), N° 6.
- González M., Pilar (1989). *"Aplicación del LISREL al análisis del rendimiento estudiantil"*. Economía (Mérida, Venezuela), N° 4.
- González S., Humberto; Mora S., Leonor y Villegas, Carolina (2002). *"Desempeño estudiantil y equidad en la universidad venezolana"*. Cuadernos OPSU N° 6. Proyecto Alma Mater, Caracas, Venezuela.
- McCullagh, Peter (1980) *"Regression model for ordinal data"*. Journal of the Royal Statistical Society, B series, 42, N° 2, pp. 109-142.
- Porto, Alberto y Di Gresia, Luciano (2001). *"Rendimiento de estudiantes universitarios y sus determinantes"*. Consultado el 04-Marzo-2008 en <http://www.depeco.econo.unlp.edu.ar/espec/espec2.pdf>.
- Rodríguez, Germán (2008) *"Lecture Notes: Generalized Linear Model Theory"*. Princeton University. Consultado el 26-Mayo-2008 en <http://data.princeton.edu/wws509/notes/a2.pdf>
- Stokes, Maura; Davis, Charles y Koch, Gary (2000). *"Categorical data analysis using the SAS® system"*. 2ª Edición. North Carolina: SAS Institute Inc.
- Varela, J., Sinha, S., Ponsot, E. y Valera, J (2008). *"Valor pronóstico del k-ésimo periodo inicial sobre rendimiento de los estudiantes de la FACES-ULA."* Sometido a arbitraje en Actualidad Contable FACES.