Valor pronóstico del k-ésimo período académico en el rendimiento de los alumnos de la FACES-ULA[1]

Varela, José Leonardo; Sinha, Surendra; Ponsot Balaguer, Ernesto; Valera, Jorge

Varela, José Leonardo
Licenciado en Matemáticas.
M.Sc. en Matemáticas
Profesor Asistente en el Instituto Universitario de Tecnológica de Cumaná.
leovarela01@hotmail.com,
leovarela01@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/

Ponsot Balaguer, 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

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

Recibido: 12-06-2008
Revisado: 19-01-2009
Aceptado: 27-03-2009

Se presenta un modelo de regresión logística (el modelo de posibilidad proporcional). El mismo permite establecer si factores como Edad y promedio en los k=1,..,4 períodos académicos de la carrera explican el rendimiento en el resto de la carrera y de igual manera interpretar los resultados en términos de posibilidad. El rendimiento se mide en base al promedio de calificaciones. La muestra consta de 985 estudiantes de la Facultad de Ciencias Económicas y Sociales (FACES) de la Universidad de Los Andes (ULA), con ingreso a partir del año 2003. Se probó estadísticamente que el período académico que mejor pronostica el rendimiento estudiantil en cada una de las carreras es el primero. La Edad contribuye a explicar el rendimiento estudiantil sólo en la carrera de Contaduría no siendo así para las otras carreras. Se encontró que hay diferencias significativas en el rendimiento para cada una de las carreras de dicha Facultad.
Palabras clave: Regresión logística, modelo de posibilidad proporcional, rendimiento estudiantil.

RESUMEN

A logistic regression model (the proportional odds model) is presented. It enables to determine whether factors such as Age and the grade point average in the k=1,…,4 career academic periods explain the performance in the rest of the career and similarly to interpret the results in terms of possibility. The performance is measured based on the average grades. The sample consists of 985 students from the Faculty of Economics and Social Sciences (FACES) of the University of Los Andes (ULA), with admission since the year 2003. It is statistically established that the academic period that best predicted the student achievement in each of the career is the first. The Age helps explain student achievement only in the career of Accountancy but not in the other ones. It was found that there are significant differences in the performance for each of the careers in that Faculty.
Keywords: Logistic regression, proportional odds model, student achievement.

ABSTRACT

[1] Este trabajo fue elaborado en el marco del “Seminario de Modelos Lineales Generalizados” del Doctorado en Estadística, IEAC-FACES-ULA, conducido por el profesor Surendra Sinha

1. Introducción

El bajo rendimiento estudiantil es uno de los problemas más graves que actualmente tiene el sistema universitario en Venezuela. De ahí que, el conocimiento de las causas que hacen que el rendimiento sea bajo permitiría implementar políticas para mejorarlo.

Según la literatura revisada, el problema del rendimiento estudiantil se ha atacado básicamente desde dos enfoques: el primero se basa en la presentación de estadísticas descriptivas (Maldonado y Álvarez, 2005), mientras que el segundo se basa en la aplicación de la estadística clásica.

En la Universidad de Los Andes (ULA), algunos autores han estudiado el problema del rendimiento estudiantil dentro del segundo de los enfoques. González (1989) y Garnica et al (1991), abordaron el tema basados en modelos LISREL y Análisis Discriminante, respectivamente. Más recientemente, Reyes (2001), basó su estudio en un análisis estadístico multivariante. En tanto que Ponsot et al (2008) y Valera et al (2008), basan sus estudios en la metodología de regresión logística pero indagando sobre objetivos distintos a los de esta investigación.

En este trabajo se estudia el problema del rendimiento estudiantil buscando un valor pronóstico de la carrera basado en el menor número de períodos académicos cursado por los estudiantes y utilizando para ello la metodología de regresión logística. La misma se basa en el modelo de posibilidad proporcional y en el cálculo de logits acumulados, y permite de esta manera interpretar los resultados en términos de posibilidades, respaldando los resultados a través de la inferencia clásica. El estudio se limitó únicamente a los estudiantes de FACES, sin embargo, la metodología utilizada puede aplicarse a cualquier facultad dentro de la ULA o alguna otra institución educativa.

Objetivo general

Encontrar medidas de pronóstico sobre el rendimiento estudiantil en la FACES-ULA.

Objetivos específicos

1. Determinar el número de períodos académicos iniciales requerido por carrera (Administración, Contaduría, Economía y Estadística), para establecer el mejor pronóstico del rendimiento estudiantil.
2. Aplicar la metodología de regresión logística para pronosticar el rendimiento estudiantil.

2. Metodología utilizada.

2.1 Descripción de los datos.

El presente estudio se basa en datos de estudiantes de las carreras de Administración (*Adm*), Contaduría (*Cont*), Economía (*Eco*) y Estadística (*Estad*) de la ULA. Los datos se obtuvieron gracias a la colaboración de la Dirección de Servicios de Información Administrativa de la ULA, en la persona de su coordinador, y constituyen la fuente oficial de la información sobre inscripciones estudiantiles semestre a semestre.

La población de estudiantes va desde el año 1974 hasta el año 2007. La base de datos contiene información acerca de 10.584 alumnos de la Facultad de ciencias Económicas y Sociales (FACES). No obstante, en conversaciones sostenidas con quienes facilitaron los datos, se ha concluido que la máxima confiabilidad se obtendría en los datos correspondientes a los alumnos cuyo ingreso haya ocurrido desde el año 2003 al 2007. Por tal motivo, se consideró ese período de tiempo como base para realizar el estudio. El cuadro 1 muestra el número de estudiantes registrados en FACES por carreras desde el año 2003 hasta el año 2007.

Cuadro 1
Alumnos ingresados en los años 2003-2007 en la FACES-ULA

	Carreras				
Año de ingreso	*Adm*	*Cont*	*Eco*	*Estad*	**Total**
2003	74	84	56	27	**241**
2004	154	165	79	50	**448**
2005	184	197	102	54	**537**
2006	242	225	121	44	**632**
2007	142	145	81	26	**394**
Total	**796**	**816**	**439**	**201**	**2252**

Fuente: Elaboración propia

Para fines del estudio, esto deja en principio un total de 2..252 alumnos. Los datos consisten básicamente de edades y promedios de notas de cada estudiante. Además, como criterio de selección, de los 2.252 se consideraron sólo los estudiantes con un mínimo de 5 períodos cursados dentro de la carrera, esto para hacerlos comparables, entendiéndose como período cursado el lapso durante el cual el estudiante cursa asignaturas (Semestres ó mini-semestres). Quedan para el estudio un total de 985 estudiantes, de los cuales 345 corresponden a la carrera de *Adm*, 403 a la carrera de *Cont*, 184 a *Eco* y 53 a *Estad*. Puesto que el tamaño de muestra en las carreras de *Eco* y *Estad* no son muy grandes (lo cual puede presentar problemas de convergencia en los modelos utilizados), se consideraron dichas carreras como una sola.

Se cuenta con las siguientes variables: Carrera, Edad, $PromP_1$, $PromPR_1$, $PromP_2$, $PromPR_2$, $PromP_3$, $PromPR_3$, $PromP_4$, $PromPR_4$. La tabla 2 muestra la descripción de cada una de las variables, el nivel de referencia y la categoría definida en uno de los niveles. La variable $PromP_1$ se refiere al promedio del primer período cursado por cada uno de los estudiantes, mientras que la variable $PromPR_1$ se refiere al promedio del resto de los períodos cursados por cada uno de los estudiantes. Por ejemplo, supongamos que un estudiante cualquiera ha cursado 6 períodos en la carrera de *Estad*, entonces $PromP_1 = 15{,}2$ implica que el estudiante obtuvo un promedio de calificaciones de 15,2 puntos durante el primer período de asignaturas cursado. $PromPR_1 = 16{,}1$ implica que el estudiante obtuvo un promedio de calificaciones de 16,1 puntos correspondiente a las asignaturas cursadas en los períodos 2, 3 ,4 , 5 y 6. Por otra parte, la variable $PromP_2$ se refiere al promedio de notas en los dos primeros períodos cursados, mientras que $PromPR_2$ se refiere al promedio de notas del tercer período cursado en adelante. De manera análoga se definen las variables $PromP_3$, $PromPR_3$, $PromP_4$ y $PromPR_4$. Es importante destacar que para el cálculo del promedio de notas no se consideraron las asignaturas retiradas legalmente por el alumno, ni aquellas cuya calificación fue asentada estando el alumno incurso en medidas de rendimiento estudiantil. También puede observarse que la variable de respuesta $PromPR_k$ (para todo k=1,..,4), es originalmente continua pero fue categorizada para este estudio. Se tomaron en cuenta tres niveles de respuesta (3, 2 y 1), indicando esto el aumento en el promedio de notas en la variable de respuesta $PromPR_k$ (para todo k=1,..,4), de *deficiente* a *bueno*. Análogamente, se realizó este procedimiento para la variable explicativa $PromP_k$ (para todo k=1,..,4).

Cuadro 2
Descripción de las variables

Variable	Descripción	Rango	Categoría
		3. Mayor o igual que 21 años	**>= 21**
Edad	Edad de ingreso a la ULA.	2. Entre 19 y 20 años	**[19;20]**
		1. Menor o igual que 18 años	**<= 18**
		3. Promedio mayor que 15 puntos	**> 15**
$PromP_k$	Promedio de notas obtenido hasta el *k*-ésimo período académico cursado por cada estudiante	2. Promedio entre 10 y 15 puntos	**[10;15]**
		1. Promedio menor que 10 puntos	**< 10**
		3. Promedio mayor que 15 puntos	***Bueno***
$PromPR_k$	Promedio de notas obtenido desde el (*k+1*)-ésimo período académico cursado por cada estudiante en adelante.	2. Promedio entre 10 y 15 puntos	***Regular***
		1. Promedio menor que 15 puntos	***deficiente***
Carrera	Carrera que cursa.	Administración Contaduría Economía o Estadística	***Adm*** ***Cont*** ***Eco-Estad***
Donde *k*=1,..,4 denota los primeros 4 períodos académicos cursados por cada estudiante			

Fuente: Elaboración propia

Luego de categorizadas las variables, se procedió a obtener las tablas de contingencia. El cuadro 3 muestra de manera resumida, para las carreras de *Administración* y *Contaduría Pública,* la cantidad de estudiantes con $PromPR_k$ *(k=1,..4) bueno*, *regular* o *deficiente* dentro de cada uno de los grupos de edad y $PromP_k$ *(k=1,..,4),* teniendo en cuenta que cuando se estén observando valores de la variable $PromPR_k$ (para algún valor de *k*), se asumirá el mismo valor de *k* para la variable $PromP_k$.

Cuadro 3
Calificaciones de estudiantes para las carreras de *Administracióm* y *Contaduría* considerando las variables Edad, $PromP_k$, $PromPR_1$, $PromPR_2$, $PromPR_3$ y $PromPR_4$.

Carrera	Edad	$PromP_k$	$PromPR_1$			$PromPR_2$			$PromPR_3$			$PromPR_4$		
			bueno	***regular***	***defi.***	***bueno***	***regular***	***defi***	***bueno***	***regular***	***Defi***	***bueno***	***regular***	*defi*
Adm	<=18	>15	19	5	0	20	5	0	24	3	1	25	5	1
	"	10-15	15	47	4	17	51	5	12	57	5	13	52	5
	"	<10	1	20	15	0	14	14	0	12	12	1	13	11
	19-20	>15	14	3	0	13	2	0	13	5	0	14	4	0
	"	10-15	5	72	5	11	78	8	14	81	5	18	80	5
	"	<10	1	42	15	0	33	12	0	25	14	1	27	8
	>=21	>15	6	5	0	5	4	0	9	2	0	9	2	0
	"	10-15	4	23	2	5	26	3	4	26	2	7	23	2
	"	<10	0	12	10	1	10	8	1	8	10	2	11	6
Cont	<=18	>15	48	1	0	47	3	0	49	3	0	50	5	0
	"	10-15	15	52	3	14	60	4	16	58	4	11	62	4
	"	<10	1	25	8	1	16	8	1	14	8	2	12	7
	19-20	>15	34	7	0	34	4	0	37	1	0	39	3	0
	"	10-15	14	50	8	20	63	6	20	63	8	20	64	8
	"	<10	1	39	15	0	27	14	1	25	13	2	17	15
	>=21	>15	14	5	0	12	6	0	11	6	0	11	6	0
	"	10-15	3	29	4	6	33	4	4	37	5	6	37	5
	"	<10	0	16	11	0	12	9	1	9	9	1	7	9
Donde *k*=1,..,4 denota los primeros 4 períodos académicos cursados por cada estudiante y *defi.=deficiente*														

Fuente: Elaboración propia

Por otro lado, el cuadro 4 muestra de manera resumida, para la carrera de *Eco-Estad,* la cantidad de estudiantes con $PromPR_k$ *(k=1,..4) bueno*, *regular* o *deficiente* dentro de cada uno de los grupos de Edad y $PromP_k$ *(k=1,..,4).* Al igual que la tabla 3, debe tenerse en cuenta que cuando se estén observando valores de la variable $PromPR_k$ (para algún valor de *k*), se asumirá el mismo valor de *k* para la variable $PromP_k$.

Cuadro 4
Calificaciones de estudiantes para las carrera *Económia y Estadística* considerando las variables Edad, Edad, $PromP_k$, $PromPR_1$, $PromPR_2$, $PromPR_3$ y $PromPR_4$.

Carrera	Edad	$PromP_k$	$PromPR_1$			$PromPR_2$			$PromPR_3$			$PromPR_4$		
			bueno	***regular***	***defi***	***bueno***	***regular***	***defi***	***bueno***	***regular***	***defi***	***bueno***	***regular***	*defi*
Eco-Estad	<=18	>15	14	6	0	16	5	0	15	4	0	17	4	0
	"	10-15	8	44	8	6	42	11	8	42	10	7	45	9
	"	<10	0	8	6	0	7	7	1	11	3	4	5	3
	19-20	>15	6	10	0	7	5	0	8	6	0	8	4	0
	"	10-15	6	56	3	7	58	7	7	55	9	5	58	8
	"	<10	0	11	9	0	9	8	3	6	7	4	8	6
	>=21	>15	3	1	0	2	1	0	4	0	0	3	1	0
	"	10-15	1	19	6	4	16	4	2	19	2	3	17	2
	"	<10	0	7	5	1	7	7	0	6	9	0	6	10
Donde *k*=1,..,4 denota los primeros 4 períodos académicos cursados por cada estudiante y *defi.=deficiente*														

Fuente: Elaboración propia

2.2 Regresión logística

La regresión logística es una técnica de modelado estadístico que es apropiada cuando se tiene una variable de respuesta del tipo categórica. La misma describe la relación entre una variable de respuesta categórica y un conjunto de variables explicativas. La variable de respuesta es frecuentemente dicotómica (dos niveles), sin embargo, puede ser politómica (más de dos niveles). Estos niveles en la variable de respuesta pueden ser de escala nominal u ordinal.

Para el caso de estudio, la variable de respuesta PromPR_k (k=1,..,4) es *politómica*, puesto que tiene tres niveles (*bueno, regular* y *deficiente*), y, además, dichos niveles son de escala ordinal. En este caso se modelan funciones llamadas *logits acumulados* por medio del *modelo de posibilidad proporcional* (McCullagh:1980).

2.3 Modelo de posibilidad proporcional

Se consideran cuatro modelos para cada carrera. Para ilustrar la manera en la que se usa el *modelo de posibilidad proporcional* tomemos los datos de la tabla 3, sólo para la carrera de *Adm*. El cuadro 5 muestra las variables a ser consideradas para cada uno de los modelos dentro de dicha carrera. Como se puede observar, el modelo1 consiste en tomar como variables explicativas la Edad y PromP_1, y como variable de respuesta PromPR_1. Para efectos ilustrativos del *modelo de posibilidad proporcional* nos quedaremos con este primer modelo.

Cuadro 5
Variables a ser consideradas en cada uno de los modelos dentro de la carrera de *Administración*.

	Variables		
Modelo	Variables explicativas		Variable de respuesta
*Adm*1	Edad	PromP_1	PromPR_1
*Adm*2	Edad	PromP_2	PromPR_2
*Adm*3	Edad	PromP_3	PromPR_3
*Adm*4	Edad	PromP_4	PromPR_4

Fuente: Elaboración propia

Asumiendo que los datos provienen de un muestreo estratificado simple, o que por lo menos son conceptualmente representativos de una población estratificada, los mismos tienen la siguiente verosimilitud:

$$\Pr(n_{hij}) = \prod_{h=1}^{3}\prod_{i=1}^{3} n_{hi+}!\prod_{j=1}^{3}\frac{\pi_{hij}^{n_{hij}}}{n_{hij}!}, \text{ donde } \sum_{j=1}^{3}\pi_{hij} = 1$$

$$\text{y } n_{hi+} = (n_{hi1} + n_{hi2} + n_{hi3})$$

con n_{hi1}, n_{hi2} y n_{hi3} son el número de personas de hth Edad y con ith promedio de notas en el primer período.

Una posible estrategia de análisis consiste en crear una variable de respuesta dicotómica combinando dos de los niveles de la respuesta y de esta manera basar un modelo sobre Pr(obtener un promedio *bueno* en el resto de la carrera) versus Pr(obtener un promedio *regular o deficiente* en el resto de la carrera) o Pr(obtener un promedio *bueno o regular* en el resto de la carrera) versus Pr(obtener un promedio *deficiente* en el resto de la carrera). Cabe destacar que como hay un orden natural para estos niveles de respuesta, tiene sentido considerar una estrategia que tome ventaja de este orden. Consideremos las cantidades,

$$\theta_{hi1} = \pi_{hi1}, \quad \theta_{hi2} = \pi_{hi1} + \pi_{hi2}, \qquad (1)$$

donde π_{hi1} denota la probabilidad de obtener un promedio *bueno* en el resto de la carrera, π_{hi2} denota la probabilidad de obtener un promedio *regular* en el resto de la carrera y π_{hi3} denota la probabilidad de obtener un promedio *deficiente* en el resto de la carrera. En tanto que, θ_{hij} representa las probabilidades acumuladas, es decir: θ_{hi1} es la probabilidad de obtener un promedio *bueno* en el resto de la carrera, y θ_{hi2} es la probabilidad de obtener un promedio *bueno* o *regular* en el resto de la carrera (según los niveles *h*-ésimo y *i*-ésimo de las variables Edad y PromP_k respectivamente).

Para una respuesta dicotómica se calcula una función logit para cada subpoblación, no obstante, para una respuesta *politómica* se genera más de un logit para cada subpoblación. En este caso, como se tienen datos ordinales se calculan logits acumulados, los cuales están basados en probabilidades acumuladas. Para tres niveles de respuesta se calculan dos logits acumulados:

$$\text{logit}(\theta_{hi1}) = \log\left[\frac{\theta_{hi1}}{1-\theta_{hi1}}\right] = \log\left[\frac{\pi_{hi1}}{1-\pi_{hi1}}\right] = \log\left[\frac{\pi_{hi1}}{\pi_{hi2}+\pi_{hi3}}\right] \text{ y}$$

$$\text{logit}(\theta_{hi2}) = \log\left[\frac{\theta_{hi2}}{1-\theta_{hi2}}\right] = \log\left[\frac{\pi_{hi1}+\pi_{hi2}}{1-\pi_{hi1}-\pi_{hi2}}\right] = \log\left[\frac{\pi_{hi1}+\pi_{hi2}}{\pi_{hi3}}\right]$$

Estos logits acumulados son los logaritmos de posibilidad de obtener un promedio *bueno* a *regular* o *deficiente* en el resto de la carrera y los logaritmos de posibilidad de obtener un promedio *bueno* o *regular* a *deficiente* en el resto de la carrera respectivamente. Ambos logaritmos de posibilidad parten de lo más favorable a lo menos favorable. El *modelo de posibilidad proporcional* toma ambos casos en cuenta.

El modelo que aplica para ambos logits de manera simultánea para cada combinación de Edad y PromP_1 es,

$$\text{logit}(\theta_{hik}) = \alpha_k + \mathbf{x}'_{hi}\beta_k$$

donde *k* indexa los dos logits, $\mathbf{x}_{hi}$ es el vector en la matriz de diseño correspondiente al logit y β_k el vector de parámetros para cada logit. Esto dice que hay parámetros de intercepto separados α_k y diferentes conjuntos de parámetros de regresión β_k para cada logit.

Al tomar diferencias en los logits entre dos sub-poblaciones para este modelo, se obtiene,

$$\text{logit}(\theta_{hik}) - \text{logit}(\theta_{hi'k}) = (\mathbf{x}_{hi} - \mathbf{x}_{hi'})'\beta_k \text{ para } k = 1,2.$$

de manera que es necesario mirar las diferencias en los logits simultáneamente para comparar la respuesta entre dos sub-poblaciones. El supuesto para que se pueda aplicar el *modelo de posibilidad proporcional* es que $\beta_k = \beta$ para todo *k,* simplificando el modelo como

$$\text{logit}(\theta_{hik}) = \alpha_k + \mathbf{x}'_{hi}\beta \quad .$$

Este modelo puede también establecerse como

$$\theta_{hik} = \frac{\exp(\alpha_k + \mathbf{x}'_{hi}\beta)}{1+\exp(\alpha_k + \mathbf{x}'_{hi}\beta)} \qquad (2)$$

En notación matricial el modelo puede expresarse como en la tabla 6. Este modelo corresponde a uno de efectos principales, y la parametrización es del tipo incremental, donde los parámetros α_1 y α_2 son los interceptos para el primero y segundo logit acumulado respectivamente, β_1 es el efecto incremental para Edad de 19 a 20 años, β_2 es el efecto incremental para Edad >=21 años, β_3 es el efecto incremental para $PromP_1$ entre 10 y 15 puntos, y β_4 es el efecto incremental para $PromP_1$ >15 puntos. Como los estudiantes con Edad <=18 años y con $PromP_1$ < 10 puntos son descritos por los interceptos α_1 y α_2 para el primero y segundo logit acumulado respectivamente, estas celdas se conocen como celdas de referencia.

Cuadro 6
Modelo de efectos principales para el logit en notación matricial

$$\begin{bmatrix} \text{logit}(\theta_{111}) \\ \text{logit}(\theta_{112}) \\ \text{logit}(\theta_{121}) \\ \text{logit}(\theta_{122}) \\ \text{logit}(\theta_{131}) \\ \text{logit}(\theta_{132}) \\ \text{logit}(\theta_{211}) \\ \text{logit}(\theta_{212}) \\ \text{logit}(\theta_{221}) \\ \text{logit}(\theta_{222}) \\ \text{logit}(\theta_{231}) \\ \text{logit}(\theta_{232}) \\ \text{logit}(\theta_{311}) \\ \text{logit}(\theta_{312}) \\ \text{logit}(\theta_{321}) \\ \text{logit}(\theta_{322}) \\ \text{logit}(\theta_{331}) \\ \text{logit}(\theta_{332}) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}$$

Fuente: Elaboración propia

Una vez se ha corrido el modelo, es necesario asegurar el buen ajuste de los datos. Lo primero que debe verificarse es la suposición de *posibilidad proporcional*. Existen programas que permiten verificar dicha suposición por medio de una prueba o Test de puntuación, donde la hipótesis nula es que $\beta_k = \beta$ para todo *k*, es decir, que los efectos son los mismos para cada logit acumulado. De otro lado, para asegurar el buen ajuste de los datos, las dos pruebas de bondad de ajuste más utilizados son el chi-cuadrado de Pearson y el chi-cuadrado razón de verosimilitud conocido como *deviance*.

Es importante destacar que en el caso en que se rechace la hipótesis nula en la prueba de puntuación para la suposición de posibilidad proporcional, es necesario considerar otro modelo tal como logit generalizado (ver Stokes et al, 2000:257-264).

3. Resultados

El cuadro 7 muestra tanto la prueba de puntuación para la suposición de *posibilidad proporcional* como las pruebas de bondad de ajuste (Deviance y Pearson), para cada uno de los modelos dentro de cada una de las carreras (los valores de los grados de libertad, denotados como DF, tanto para la prueba de *posibilidad proporcional* como para las pruebas de bondad de ajuste son 4 y 12 respectivamente para todos los modelos). En la misma puede observarse que con excepción del modelo *Eco-Estad p*ara la carrera de *Eco-Estad* , en todos los demás modelos se acepta la hipótesis nula para la prueba de posibilidad proporcional. Es importante aclarar que ambas pruebas de bondad de ajuste, tanto el Deviance como el de Pearson, indican que todos los modelos con excepción del modelo *Eco-Estad* 4 para la carrera de *Eco-Estad* ajustan de manera adecuada los datos.

Cuadro 7
Estadísticos de bondad de ajuste y posibilidad proporcional para todos los modelos ($\gamma = 5\%$).

			Prueba de bondad de ajuste					
	Prueba puntuación para sup. de posibilidad proporcional		Deviance			Pearson		
Modelo	ChiSq	Pr> ChiSq	valor	Valor/DF	Pr> ChiSq	valor	Valor/DF	Pr> ChiSq
*Adm*1	7.7595	0.1008	14.1371	1.1781	0.2920	13.3342	1.1112	0.3452
*Adm*2	5.7821	0.2160	15.494	1.2912	0.2155	14.17	1.1812	0.2897
*Adm*3	8.5473	0.0735	11.2066	0.9339	0.5113	19.22	1.6022	0.0832
*Adm*4	6.9431	0.1389	8.2010	0.6834	0.7692	11.25	0.9383	0.5068
*Cont*1	2.9578	0.5649	9.895	0.8246	0.6252	8.6067	0.7172	0.7361
*Cont*2	2.0422	0.7280	8.4733	0.7061	0.7471	7.0522	0.5877	0.8541
*Cont*3	2.9280	0.5699	9.0310	0.7526	0.7003	9.8194	0.8183	0.6318
*Cont*4	8.4751	0.0756	11.811	0.9843	0.4609	13.1705	1.0975	0.3568
Eco-Estad 1	5.9294	0.2045	13.4882	1.1240	0.3346	12.2323	1.0194	0.4272
Eco-Estad 2	3.8064	0.4328	7.1378	0.5948	0.8484	6.6864	0.5572	0.877
Eco-Estad 3	4.5722	0.3341	18.5758	1.5480	0.0993	21.6930	1.8078	0.04
Eco-Estad 4	24.3737	<.0001	26.5138	2.2095	0.0091	33.7976	2.8165	0.0007

Fuente: Elaboración propia

Teniendo en cuenta que lo que se busca es conseguir un valor pronóstico de la carrera basado en el menor número de períodos cursado por los estudiantes, el criterio utilizado para escoger el modelo que mejor ajuste los datos consiste en tomar para cada carrera aquel modelo que ajuste de manera satisfactoria los datos y además esté basado en el menor número de períodos académicos cursados por los estudiantes. En la tabla 7, para la carrera de *Adm* cualquiera de los modelos ajusta de manera satisfactoria los datos, de manera que basado en el criterio de escogencia del modelo se tiene que el modelo *Adm*1 es el que mejor ajusta los datos y de aquí que el primer período académico cursado por los estudiantes sirve como pronóstico del rendimiento estudiantil en el resto de la carrera. El mismo criterio se usa para las carreras de *Cont* y *Eco-Estad*, y, al igual que para *Adm,* el primer período académico sirve como valor pronóstico del rendimiento estudiantil en ambas carreras.

El cuadro 8 muestra la significancia de los efectos para cada uno de los modelos que mejor pronostican el rendimiento estudiantil, docimando la hipótesis de nulidad de los efectos presentes en cada uno de los modelos. Para ello fue usada la prueba de Wald.

Cuadro 8
Análisis para los efectos principales($\alpha = 5\%$).

Análisis de efectos						
	Edad			$PromP_1$		
Modelo	DF	ChiSq Wald	Pr>ChiSq	DF	ChiSq Wald	Pr>ChiSq
*Adm*1	2	2.9750	0.2259	2	109.7954	<.0001
*Cont*1	2	11.7942	0.0027	2	153.4125	<.0001
Eco-Estad 1	2	1.9935	0.3691	2	66.2534	<.0001

Fuente: Elaboración propia

A un nivel del 5%, se afirma que el efecto $PromP_1$ resulta significativo para cada uno de los modelos, mientras que el efecto Edad solo resulta significante para el modelo Cont1. Indicando esto que tanto la Edad como $PromP_1$ contribuyen a explicar el rendimiento en la carrera de Cont, mientras que para las carreras de Adm y Eco-Estad el factor que contribuye a explicar el rendimiento en la carrera es $PromP_1$.

El cuadro 9 contiene tanto la estimación de los parámetros como de las posibilidades (sólo para los efectos que resultaron significantes).

Cuadro 9
Estimación de parámetros y posibilidades ($\alpha = 5\%$).

Carrera	Parámetro		DF	Estim.	Error stand	ChiSq Wald	Pr>ChiSq	Estim. de Posib.	95% Wald Lím. De conf.	
	intercepto	*bueno*	1	-3.78	0.39	91.04	<.0001			
	intercepto	*regular*	1	0.85	0.26	10.12	0.0015			
Adm.	$PromP_1$	10-15	1	2.09	0.33	38.06	<.0001	8.1	4.172	15.7
	$PromP_1$	>15	1	5.09	0.48	109.74	<.0001	164.02	62.79	422.2
	intercepto	*bueno*	1	-2.89	0.35	68.12	<.0001			
	intercepto	*regular*	1	1.37	0.28	23.76	<.0001			
	Edad	19-20	1	-0.54	0.28	3.74	0.0529	0.582	0.336	1.007
Cont	Edad	>=21	1	-1.14	0.33	11.71	0.0006	0.319	0.166	0.61
	$PromP_1$	10-15	1	1.69	0.31	29.74	<.0001	5.443	2.961	10
	$PromP_1$	>15	1	5.36	0.44	149.05	<.0001	214.036	90.440	506.5
	intercepto	*bueno*	1	-3.97	0.48	68.12	<.0001			
	intercepto	*regular*	1	0.47	0.37	1.63	0.2012			
Eco-Estad	$PromP_1$	10-15	1	1.86	0.38	23.01	<.0001	6.451	3.012	13.82
	$PromP_1$	>15	1	4.43	0.54	65.93	<.0001	83.744	28.759	244

Fuente: Elaboración propia

La interpretación para las estimaciones de las posibilidades son las siguientes:

1. Un estudiante en la carrera de *Adm* con un promedio en el primer período entre 10-15 puntos, tiene $e^{2.09} \approx 8.1$ veces más posibilidad de obtener un promedio *bueno* versus *regular* o *deficiente* (igualmente para *bueno* o *regular* versus *deficiente*), en el resto de la carrera que un estudiante con un promedio en el primer período <10.
2. Un estudiante en la carrera de *Adm* con un promedio en el primer período >15 puntos, tiene $e^{5.09} \approx 164$ veces más posibilidad de obtener un promedio *bueno* versus *regular* o *deficiente* (igualmente para *bueno* o *regular* versus *deficiente*), en el resto de la carrera que un estudiante con un promedio en el primer período <10.
3. Un estudiante en la carrera de *Cont* con Edad de ingreso entre 19-20 años, tiene 1 / 0.582 =1 .72 veces más posibilidad de obtener un promedio *deficiente* versus *regular* o *bueno* (igualmente para *deficiente* o *regular* versus *bueno*), en el resto de la carrera que un estudiante con Edad de ingreso <=18.
4. Un estudiante en la carrera de *Cont* con Edad de ingreso >= 21 años, tiene 1 / 0.319 =3.13 veces más posibilidad de obtener un promedio *deficiente* versus *regular* o *bueno* (igualmente para *deficiente* o *regular* versus *bueno*), en el resto de la carrera que un estudiante con Edad de ingreso <=18.
5. Un estudiante en la carrera de *Cont* con un promedio en el primer período entre 10-15 puntos, tiene 5.443 veces más posibilidad de obtener un promedio *bueno* versus *regular* o *deficiente* (igualmente para *bueno* o

regular versus *deficiente*), en el resto de la carrera que un estudiante con un promedio en el primer período <10.

6. Un estudiante en la carrera de *Cont* con un promedio en el primer período >15 puntos, tiene 214 veces más posibilidad de obtener un promedio *bueno* versus *regular* o *deficiente* (igualmente para *bueno* o *regular* versus *deficiente*), en el resto de la carrera que un estudiante con un promedio en el primer período <10.
7. Un estudiante en la carrera de *Eco-Estad* con un promedio en el primer período entre 10-15 puntos, tiene 6.45 veces más posibilidad de obtener un promedio *bueno* versus *regular* o *deficiente* (igualmente para *bueno* o *regular* versus *deficiente*), en el resto de la carrera que un estudiante con un promedio en el primer período <10.
8. Un estudiante en la carrera de *Eco-Estad* con un promedio en el primer período >15 puntos, tiene 83 veces más posibilidad de obtener un promedio *bueno* versus *regular* o *deficiente* (igualmente para *bueno* o *regular* versus *deficiente*), en el resto de la carrera que un estudiante con un promedio en el primer período <10.

Por último, para ver si existen diferencias significativas entre las carreras, se hizo un modelo (asumiendo que el período académico que mejor pronostica el rendimiento estudiantil es el primero), donde se tomaron en cuenta como variables explicativas Carrera y $PromP_1$, y como variable de respuesta $PromPR_1$. En este caso la prueba de puntuación para la suposición de posibilidad proporcional da un valor-*p* de 0.2479 con 4 grados de libertad y un valor de *ChiSq = 5.4085,* indicando esto que se acepta la hipótesis de posibilidad proporcional. Por otro parte, la prueba Deviance con 12 grados de libertad da una probabilidad de 0.4527 con un valor de 11.9124, indicando esto que el modelo ajusta perfectamente los datos, y además valor/DF=0.9927 lo que indica que no hay problema de subdispersión. La prueba de Wald para el análisis de los efectos con 2 grados de libertad arrojó los valores *ChiSq de Wald* 18.4 y 348.13 con una probabilidad de 0.0001 y < 0.001 para los efectos Carrera y $PromP_1$ respectivamente, indicando esto que tanto la Carrera como $PromP_1$ contribuyen a explicar el rendimiento de los alumnos en FACES. Los resultados correspondientes a la estimación de los parámetros y posibilidades pueden observarse en el cuadro 10.

Cuadro 10
Estimación de parámetros y posibilidades considerando la Carrera como factor($\alpha = 5\%$).

Parámetro		DF	Estim.	Error stand	ChiSq Wald	Pr>ChiSq	Estim. de Posib.	95% Wald Lím. De conf.	
intercepto	*bueno*	1	-4.31	0.27	257.43	<.0001			
intercepto	*regular*	1	0.08	0.19	0.21	0.6424			
Carrera	*Adm*	1	0.53	0.19	7.25	0.0071	1.7	1.15	2.5
Carrera	*Cont*	1	0.83	0.19	18.39	<.0001	2.3	1.57	3.35
$PromP_1$	10-15	1	1.94	0.19	98.41	<.0001	7	4.7	10.2
$PromP_1$	>15	1	5.06	0.27	343.7	<.0001	158	92.5	270

Fuente: Elaboración propia

La interpretación para las estimaciones de posibilidad en este caso son las siguientes:

1. Un estudiante de FACES que curse la carrera de *Adm* tiene 1.7 veces más posibilidad de obtener un promedio *bueno* versus *regular* o *deficiente* (igualmente para *bueno* o *regular* versus *deficiente*), en el resto de la carrera que un estudiante que curse la carrera *Eco-Estad*.
2. Un estudiante de FACES que curse la carrera de *Cont* tiene 2.3 veces más posibilidad de obtener un promedio *bueno* versus *regular* o *deficiente* (igualmente para *bueno* o *regular* versus *deficiente*), en el resto de la carrera que un estudiante que curse la Carrera *Eco-Estad*.
3. Un estudiante de *FACES* con un promedio en el primer período entre 10-15 puntos, tiene 7 veces más posibilidad de obtener un promedio *bueno* versus *regular* o *deficiente* (igualmente para *bueno* o *regular* versus *deficiente*), en el resto de la carrera que un estudiante con un promedio en el primer período <10.
4. Un estudiante de FACES con un promedio en el primer período >15 puntos, tiene 158 veces más posibilidad de obtener un promedio *bueno* versus *regular* o *deficiente* (igualmente para *bueno* o *regular* versus *deficiente*), en el resto de la carrera que un estudiante con un promedio en el primer período <10.

4. Conclusiones

Al analizar los resultados obtenidos luego de aplicar cada uno de los modelos al caso de estudio se concluyó lo siguiente:

1. El modelo de posibilidad proporcional aplicado dentro de la metodología de regresión logística, sirve para explicar perfectamente el rendimiento estudiantil dentro de cada una de las carreras (*Adm, Cont y Eco-Estad*).
2. Aunque cualquiera de los cuatro primeros períodos académicos sirven como pronóstico del rendimiento estudiantil para cada una de las carreras (con excepción del cuarto período académico para la carrera *Eco-Estad*), se tiene que el período académico que mejor pronostica el rendimiento estudiantil dentro de cada una de las carreras es el primero.
3. Teniendo en cuenta la conclusión número 2, las variables Edad y $PromP_1$ contribuyen a explicar el rendimiento en la carrera de *Cont*, mientras que para las carreras de *Adm* y *Eco-Estad* la variable que contribuye a explicar el rendimiento es $PromP_1$.
4. El análisis de los resultados de la tabla número 9, específicamente los puntos 2, 6, y 8, afirman que para todas las carreras, un estudiante con rendimiento mayor a 15 puntos en el primer período académico, tiene una posibilidad muy elevada de obtener un rendimiento bueno en la carrera.
5. Al considerar un modelo tomando en cuenta la carrera como variable explicativa y $PromP_1$, se tiene que la variable Carrera contribuye a explicar el rendimiento estudiantil, y el análisis de los resultados indica que tienen más posibilidades de tener un rendimiento *bueno* versus *regular* o *deficiente* los estudiantes de *Cont* que los de *Adm* respecto a los de *Eco-Estad*. De igual manera, se reafirma la conclusión número 4, puesto que en la interpretación para las estimaciones de posibilidad de

la tabla 10 se observa que un estudiante de FACES con un promedio en el primer período >15 puntos, tiene mucha más posibilidad de obtener un promedio *bueno* versus *regular* o *deficiente* (igualmente para *bueno* o *regular* versus *deficiente*), en el resto de la carrera que un estudiante con un promedio en el primer período <10. Esto sugiere que los esfuerzos realizados por las autoridades con el fin de mejorar el rendimiento estudiantil dentro de la FACES, debe centrarse en el primer período académico cursado por los estudiantes.

5. Referencias

Garnica, E.; González, P.; Díaz, A. y Torres, E. (1991). *"Análisis discriminante. Estudio del rendimiento estudiantil". Revista Economía,* Nº 6, Mérida Venezuela.

González M., P. (1989). *"Aplicación del LISREL al análisis del rendimiento estudiantil". Revista Economía*, Nº 4, Mérida Venezuela.

Maldonado, E. y Álvarez, A. (1995). *Diagnóstico del rendimiento estudiantil en la Escuela de Economía (1983-1989).* Mérida Venezuela: Facultad de Ciencias Económicas y Sociales, ULA.

McCullagh, P. (1980). *Regression Models for Ordinal Data. J. R. Statist. Soc.*. Vol. 42, Nº 2, (109-142).

Ponsot, E.; Sinha, S.; Varela J. y Valera J. (2008). *Un Modelo de Regresión Logística en los Estudios Universitarios: Caso FACES ULA.* Sometido a arbitraje en la revista Actualidad Contable FACES, Facultad de Ciencias Económicas y Sociales, ULA Mérida Venezuela.

Reyes, V. (2001). *Estudio del rendimiento estudiantil de FACES-ULA, a través de un análisis estadístico multivariante.* Mérida Venezuela: Universidad de Los Andes.

Stokes, M. E.; Charles S., D. y Gary G., K. (2000). *Categorical Data Analisis Using The SAS System, Second Edition.* Cary, NC: SAS Institute Inc.

Valera, J.; Sinha, S.; Varela, J. y Ponsot, E. (2008). *Modelo de regresión logística: una aplicación en la predicción del rendimiento estudiantil universitario.* Sometido a arbitraje en la revista Ciencia e Ingeniería, Facultad de Ingeniería, ULA Mérida Venezuela.