

Un modelo logit alternativo para explicar los factores que influyen en la probabilidad de éxito en Matemáticas Empresariales

Dávila Cárdenes, Nancy (ndavila@dmc.ulpgc.es)

Dpto. de Mét. Cuant. en Economía y Gestión

Univ. de Las Palmas de Gran Canaria

García Artiles, María Dolores, (mdgartiles@dmc.ulpgc.es)

Dpto. de Mét. Cuant. en Economía y Gestión

Univ. de Las Palmas de Gran Canaria

Gómez Déniz, Emilio (egomez@dmc.ulpgc.es)

Dpto. de Mét. Cuant. en Economía y Gestión

Univ. de Las Palmas de Gran Canaria

Pérez Sánchez, José M. (josemag@ugr.es)

Dpto. de Mét. Cuantitativos

Univ. de Granada

RESUMEN

En este trabajo se pretende explicar la probabilidad de éxito o fracaso en la asignatura Matemáticas Empresariales correspondiente al Grado en Administración y Dirección de Empresas en la Universidad de Las Palmas de Gran Canaria. Para ello se intentarán detectar los factores o covariables que influyen en dicha probabilidad utilizando, primero un modelo de regresión logística clásico y, en segundo lugar, un modelo de regresión logística asimétrico. En los modelos econométricos clásicos de regresión binaria tradicionalmente se emplea la regresión logística, que se basa en el enlace simétrico de uso de la distribución

logística. El propósito de este trabajo, y a la luz de los resultados obtenidos en la muestra utilizada, es incorporar en el estudio un enlace de naturaleza asimétrica que trate de dar una mejor explicación de la variable que se desea estudiar.

Palabras clave: Bayes; Modelo logit; Modelo logit asimétrico; Matemáticas Empresariales; Evaluación

Área temática: [A5]

ABSTRACT

This paper aims to explain the probability of success or failure in the course Mathematics for Business in the Degree in Business Administration in the University of Las Palmas de Gran Canaria. For this attempt, we try to identify factors or covariates influencing this probability using first a Classical logistic regression model, and secondly, an logistic asymmetric regression model. In the classical econometric regression models traditionally binary logistic regression is used, which is based on the use of symmetrical distribution logistics. The purpose of this work, and considering the results in the sample used in the study, is to incorporate an asymmetric link to try to give a better explanation of the variable to be studied.

Key words: Bayes; Logit Model; Asymmetric Logit Model; Business Mathematics; Assesment

1 INTRODUCCIÓN

Con la implantación de los nuevos planes de estudio en el curso 2009-2010 en la Universidad de Las Palmas de Gran Canaria (ULPGC) se han incorporado modificaciones en la metodología docente y la evaluación en la mayor parte de las asignaturas. En el caso de la asignatura de Matemáticas Empresariales (ME), en cuanto a la metodología docente, además del uso de material de apoyo en la plataforma virtual, que ya se venía utilizando en los planes anteriores, los estudiantes disponen de grabaciones de vídeo en la plataforma PROMETEO que es que la ULPGC ha habilitado para ello. De este modo los estudiantes, además de las clases presenciales de teoría y prácticas, disponen de casi todo el material de la asignatura en este formato. Respecto a la evaluación, los cambios más destacables se han desarrollado con la evaluación continua, que se desarrolla a través de pruebas cuya valoración se entiende que suponen un incentivo al trabajo diario del estudiante para seguir la asignatura. Siguiendo la línea de trabajos que han abordado el análisis de factores que afectan al rendimiento académico Rúa et al. (2010), Fedriani et al.(2011), en este trabajo, y tras los resultado obtenidos en un primer trabajo Dávila et al.(2011), pretendemos avanzar un poco más en el estudio de los factores que puedan explicar la probabilidad de superar la asignatura de ME, considerando para ello entre las variables a analizar el trabajo personal del alumno, piedra angular sobre la que se define el crédito ECTS. Analizar la idoneidad del sistema ECTS desde el punto de vista de la mejora del rendimiento académico y cómo influyen las características personales académicas y organizativas del alumno ya fueron estudiadas por Florido et al.(2011). En este caso, y tras la elaboración de un cuestionario que recogía información sobre la formación previa, sobre la asignatura, sobre el material y evaluación de la asignatura y finalmente sobre el trabajo personal, y que se pasó a los estudiantes una vez finalizado el primer semestre y concluída la fase de evaluación,

se elaboró el trabajo que a continuación presentamos.

Resulta bien conocido la utilidad que los modelos de comportamiento de elección discreta aleatoria de la literatura econométrica tienen en numerosos y diferentes escenarios científicos. La regresión binaria se aplica, por ejemplo, para evaluar los riesgos que asume una entidad financiera a la hora de otorgar o no un determinado préstamo, para evaluar la posibilidad de establecer o no un determinado servicio de atención sanitaria, etc. Entre los modelos de elección discreta aleatoria, la regresión binaria, usando la regresión logística, es el procedimiento habitualmente utilizado.

En este trabajo se pretende explicar la probabilidad de éxito o fracaso en la asignatura ME de la titulación Grado en Administración y Dirección de Empresas (GADE) en la ULPGC. Para ello se intentarán detectar los factores o covariables que influyen en dicha probabilidad utilizando, primero un modelo de regresión logística (clásico y Bayesiano) y, en segundo lugar, un modelo de regresión logística Bayesiano asimétrico. En los modelos econométricos clásicos de regresión binaria tradicionalmente se emplea la regresión logística, que se basa en el enlace simétrico de uso de la distribución logística Albert y Chib (1993,1995). Sin embargo, cuando hay presencia predominante de uno de los valores de la variable endógena, los enlaces simétricos son inadecuados. Nagler (1994) indica que cuando se usa logit se está asumiendo que un individuo con una probabilidad de 0.5 de éxito es más sensible a cambios en las variables independientes, por ejemplo, cambios en una unidad en el regresor, que alguien con un 0.3 ó 0.7 de probabilidad de éxito. Sin embargo, el mismo autor sostiene que este no es necesariamente el caso en otras situaciones, donde un individuo con una probabilidad del 0.4 de éxito puede ser más sensible ante un cambio de una unidad en un regresor que otro que tiene una probabilidad del 0.5 de éxito. Si es así, la distribución es sesgada. Las respuestas no son simétricas alrededor de 0.5, por esta razón el empleo de los enlaces asimétricos es justificado en estos casos.

El propósito de este trabajo, y con los porcentajes obtenidos en la muestra

utilizada para las dos respuestas de la variable endógena del modelo, es incorporar en el estudio un enlace de naturaleza asimétrica que trate de dar una mejor explicación de la variable bajo estudio.

Este tipo de modelización ha sido aplicado recientemente en diversos trabajos. Algunos de ellos son Bermúdez et al. (2008), que ajustaron un modelo de regresión logística asimétrico para modelizar la conducta fraudulenta de los clientes de una cartera de seguros y Sáez et al. (2010), quienes utilizaron un modelo de similares características para predecir el número de infecciones en un área de cirugía de un hospital.

Determinados enlaces asimétricos para modelos de regresión binaria han sido propuestos considerando situaciones donde este enlace es más apropiado que el simétrico, sin embargo, no existe un esquema general que caracterice las formas de construcción de enlaces. El esquema que se ha seguido en este trabajo consiste en comparar la estimación Bayesiana no informativa de un modelo logit con la estimación frecuentista (clásica) del modelo logit, con el objeto de verificar que ambos modelos ofrecen resultados similares en términos de la estimación de parámetros. Posteriormente, se compararán estos resultados con la estimación Bayesiana asimétrica del modelo logit comprobándose que esta estimación Bayesiana asimétrica recoge el evidente efecto de asimetría que existe en los datos, además de mejorar notablemente el ajuste de la probabilidad de aprobar. Por último, se analizarán también las evaluaciones de todos los modelos. Se utilizará el programa WinBUGS, que permite estimar los parámetros del modelo utilizando la inferencia Bayesiana.

El resto del trabajo está organizado de la siguiente manera. En la sección 2 se explicará con cierto detalle la especificación de los modelos utilizados en este trabajo. En la sección 3 se explica la muestra utilizada así como las covariables incorporadas a los modelos analizados. La sección 4 incluye los resultados obtenidos con la aplicación de dichos modelos a la muestra de estudiantes de la asignatura ME

de GADE en la ULPGC. Finalmente, la sección 5 incluye algunas reflexiones sobre el trabajo realizado.

2 ESPECIFICACIÓN DE LOS MODELOS UTILIZADOS

En esta sección se explican los modelos de regresión logística que serán utilizados en este trabajo y que tratarán de explicar la probabilidad de éxito de un estudiante de la asignatura ME de la titulación GADE de la ULPGC mediante determinadas covariables o factores. Brevemente se explicarán el modelos logit clásico (simétrico) y una variante del mismo denominado modelo logit asimétrico. Puesto que este último se estimará utilizando la inferencia Bayesiana a través del programa WinBUGS, se señalará con cierto detalle la metodología utilizada en este último escenario.

2.1 Modelo logit frecuentista (clásico)

Partiremos de una muestra $y = (y_1, y_2, \dots, y_n)'$ que consiste de un vector de dimensión $n \times 1$ y un vector de dimensión $k \times 1$, $x_i = (x_{i1}, \dots, x_{ik})'$, de covariables o factores explicativos para cada uno de los $i = 1, \dots, n$ estudiantes que conforman la muestra. El ajuste de regresión logística (binario) clásico adaptado a nuestro estudio consistirá en estimar la probabilidad de que un estudiante supere la asignatura bajo estudio (respuesta 1) o no la supere (respuesta 0). Por tanto, si $y_i = 1$ el estudiante i supera la asignatura, mientras que $y_i = 0$ significará que no la supera. Asumimos que $y_i = 1$ con probabilidad p_i e $y_i = 0$ con probabilidad $1 - p_i$. El modelo de regresión viene dado por $p_i = F(x_i' \beta)$, donde $F(s) = 1/(1 + e^{-s})$, $-\infty < s < \infty$, es la función de distribución logística (su función inversa se denomina enlace o *link*), simétrica, esto es $F(k - s) = F(k + s)$, para una constante k y para cualquier s

real; $\beta = (\beta_1, \dots, \beta_k)'$ es un vector de dimensión $k \times 1$ formado por los parámetros a estimar, y que representa el efecto de cada una de las covariables en el modelo.

Bajo estos supuestos, la función de verosimilitud, dado los datos, viene dada por

$$\ell(y|x, \beta) = \prod_{i=1}^n [F(x'_i \beta)]^{y_i} [1 - F(x'_i \beta)]^{1-y_i}. \quad (1)$$

Los coeficientes de regresión, en este caso, pueden estimarse por el método de máxima verosimilitud, maximizando la función (1). Para ello existen numerosos paquetes estadísticos disponibles que permiten resolver este problema. Algunos de ellos son STATA, SPSS, Mathematica, etc.

2.2 Modelo logit asimétrico

Se proponen en esta sección dos modelos alternativos al modelo explicado en la sección anterior: un modelo logit Bayesiano y un modelo logit Bayesiano con carácter asimétrico. En ambos casos los coeficientes de la regresión se considerarán aleatorios eligiendo para ellos determinadas distribuciones de probabilidad (distribuciones *a priori*).

Desde el trabajo inicial de Prentice (1976) sobre la idoneidad (generando mejores ajustes) de utilizar un *link* asimétrico para el caso en que una de las respuestas sea mucho más frecuente que la otra, se han propuesto numerosos trabajos al efecto. En este sentido, cabe citar las contribuciones de Stukel (1988), Chen et al. (1999) y Bazán et al. (2006), entre otros. Este hecho ha sido confirmado en trabajos llevados a cabo en diferentes contextos. Algunos de ellos son Bermúdez et al. (2008), Chen et al. (2004), Basu y Mukhopadhyay (2000), entre otros.

Siguiendo el trabajo de Chen et al. (1999), se muestra a continuación el *link* asimétrico con el que se trabajará en esta contribución. Este modelo asume el uso

de variables auxiliares w_i , $i = 1, \dots, n$, incorporándolas de la siguiente manera:

$$y_i = \begin{cases} 1, & w_i \geq 0, \\ 0, & w_i < 0, \end{cases} \quad (2)$$

donde $w_i = x'_i\beta + \delta z_i + \varepsilon_i$, $z_i \sim G$, $\varepsilon_i \sim F$. Asumiremos independencia entre z_i y ε_i . De nuevo F representa la función de distribución logística, mientras que G es la función de distribución de la variable aleatoria half-normal (estandarizada), dada por

$$g(z) = \sqrt{\frac{2}{\pi}} e^{-z^2/2}, \quad z > 0.$$

Por otro lado, $\delta \in (-\infty, \infty)$ es un parámetro que regula el sesgo del modelo de regresión, dado por δz_i . Así, si $\delta > 0$, la probabilidad $p_i = 1$, la probabilidad de éxito del estudiante i , aumenta, mientras que si $\delta < 0$, ahora lo que se incrementa es la probabilidad de fracaso, $p_i = 0$.

Bajo este modelo, la función de verosimilitud viene dada por

$$\ell(y|x, \beta, \delta) = \prod_{i=1}^n \int_0^\infty [F(x'_i\beta + \delta z_i)]^{y_i} [1 - F(x'_i\beta + \delta z_i)]^{1-y_i} g(z_i) dz_i. \quad (3)$$

En el proceso de elección de la distribución *a priori* actuaremos de la siguiente manera. Supondremos que la distribución *a priori* de los coeficientes es normal y no informativa. Es decir, $\beta_j \sim \mathcal{N}(0, \sigma^2)$, $\forall j = 1, \dots, k$ y $\delta \sim \mathcal{N}(0, \sigma^2)$, considerando $\sigma > 0$ suficientemente grande, indicando con ello nuestro total desconocimiento *a priori* sobre los parámetros de interés y facilitando también las comparaciones con el modelo clásico. El valor que se tomará en este trabajo será $\sigma^2 = 10^{10}$.

Ahora, combinando la distribución *a priori* con la verosimilitud (3) se obtiene, vía el teorema de Bayes, la distribución *a posteriori* de los parámetros β y δ , que resulta proporcional a

$$\begin{aligned} \pi(\beta, \delta|y, x) &\propto \ell(y|x, \beta, \delta) \pi(\beta, \delta) \\ &= \left\{ \prod_{i=1}^n \int_0^\infty [F(x'_i\beta + \delta z_i)]^{y_i} [1 - F(x'_i\beta + \delta z_i)]^{1-y_i} g(z_i) dz_i \right\} \pi(\beta, \delta), \end{aligned} \quad (4)$$

donde $\pi(\beta, \delta)$ es la distribución *a priori* (bivariante) de (β, δ) .

Finalmente, y con el uso del programa WinBUGS¹ (Windows Bayesian inference Using Gibbs Sampling) se muestrea (β, δ) utilizando la distribución *a posteriori* para obtener los estimadores del modelo.

3 DESCRIPCIÓN DE LA MUESTRA

La base de datos contiene información de 279 estudiantes de ME de la titulación GADE en la ULPGC matriculados en el curso 2011–12. Las variables con las que se han realizado el estudio son ciertas características relativas al estudiante referidas a su formación previa, su opinión sobre diversos aspectos de la asignatura, así como sobre el trabajo personal del mismo.

Con respecto a su formación previa, contamos con las variables “tecnológico”, “humani”, “ciclofor”, “accesotros” y “privado”. Las 4 primeras toman el valor 1 si el alumno proviene de Bachillerato Científico Tecnológico, Humanidades, ciclo formativo u otros, respectivamente. Se ha considerado como variable de referencia, la opción de Bachillerato de Ciencias Sociales. Por otro lado, la variable “privado” toma valor 1 si el centro de estudios de secundaria fue privado.

Las variables utilizadas para valorar la opinión del alumnado sobre distintos factores de la asignatura son: “contenidos”, “interés”, “matvirtual”, “prometeo” y “videos”. La variable “contenidos” tomará el valor 1 si el alumno cree que los contenidos de la asignatura le han resultado de mucha dificultad o imposibles de seguir y 0 en otro caso. La variable “interés” es igual a 1 si al alumno le ha parecido interesante la asignatura en general y 0 en otro caso. Las restantes variables hacen

¹Programa desarrollado conjuntamente por el MRC Biostatistics Unit (University of Cambridge, Cambridge, UK) y el Imperial College School of Medicine en St. Mary’s, London) (Lunn et al. (2000)), basado en el muestreo Gibbs y aplicando técnicas de Markov Chain Monte Carlo (MCMC) (para más detalle consultar Carlin y Polson (1992) y Gilks et al. (1995)).

referencia al apoyo a través de internet de la asignatura, tomando valores 1 para una valoración positiva del material virtual (“matvirtual”); si ha utilizado la plataforma virtual PROMETEO (variable “prometeo”) y, por último, si considera que los videos de práctica colgados en la web le han resultado de utilidad (“videos”).

Por último, con respecto al trabajo personal del alumno, disponemos de las variables “extras” (igual a 1 si ha asistido a clases extras de la asignatura), “teoriatodas” (igual a 1 si ha asistido a todas las clases de teoría), “practodas” (igual a 1 si ha asistido a todas las clases de práctica), “horas2” (igual a 1 si ha estudiado menos de 2 horas al día), “considera23” (igual a 1 si considera que debería de haber estudiado entre 2 y 3 horas semanales), “tutorias5” (igual a 1 si ha asistido a tutorías más de 5 veces).

En la Tabla 1 se presenta un resumen de la codificación realizada para las variables explicativas. Además, en la Tabla 2, se presenta un resumen descriptivo de todas las variables utilizadas en este estudio. Se observa que la variable dependiente, aprobar la asignatura, tiene 183 valores en cero frente a los 96 valores en 1, por lo que se concluye que el número de estudiantes que superaron la asignatura ha sido del 34.41%.

4 RESULTADOS

Como ya se comentó anteriormente, se proponen dos alternativas de modelos de elección discreta para ajustar la probabilidad de aprobar la asignatura. Consideraremos, en primer lugar, modelos logit de estimación clásica (o frecuentista) y Bayesiana y, posteriormente, se asume un logit asimétrico desde el punto de vista Bayesiano.

Tabla 1: Variables explicativas

Variable	Definición
tecnológico	Igual a 1 si la opción de acceso a la Universidad es el Bachillerato Científico Tecnológico
humani	Igual a 1 si la opción de acceso a la Universidad es el Bachillerato de Humanidades
ciclofor	Igual a 1 si la opción de acceso a la Universidad es un Ciclo formativo
accesotros	Igual a 1 si la opción de acceso a la Universidad es Otros
privado	Igual a 1 si el centro de estudios de secundaria fue privado Igual a 0 si era público o concertado
contenidos	Igual a 1 si los contenidos de la asignatura le han resultado de mucha dificultad (o imposibles)
interés	Igual a 1 si lo estudiado en la asignatura le ha resultado de mucho interés
matvirtual	Igual a 1 si considera que el material virtual es adecuado para el seguimiento de la asignatura
prometeo	Igual a 1 si ha utilizado la plataforma PROMETEO para el seguimiento de la asignatura
videos	Igual a 1 si considera que la visualización de videos de práctica ha resultado de utilidad
extras	Igual a 1 si ha asistido a clases extras de la asignatura
teoriatodas	Igual a 1 ha asistido a todas las clases de teoría
practodas	Igual a 1 ha asistido a todas las clases de práctica
horas2	Igual a 1 ha estudiado menos de 2 horas al día Igual a 0 si no ha estudiado o ha estudiado más de 2 horas
considera23	Igual a 1 si considera que tenía que haber estudiado entre 2 y 3 horas semanales Igual a 0 si considera que tenía que haber estudiado más de 3 o menos de 2 horas
tutorias5	Igual a 1 si ha asistido a tutorías más de 5 veces

Tabla 2: Resumen descriptivo de variables

Variable	(1)	(0)
Aprobado	96 (34.41%)	183 (65.59%)
Contenidos	173 (62.00%)	106 (38.00%)
interés	132 (47.31%)	147 (52.69%)
matvirtual	146 (52.33%)	133 (47.67%)
prometeo	225 (80.64%)	54 (19.36%)
videos	219 (78.49%)	60 (21.51%)
extras	77 (27.60%)	202 (72.40%)
tecnológico	27 (9.68%)	252 (90.32%)
humani	47 (16.85%)	232 (83.15%)
ciclofor	19 (6.81%)	260 (93.19%)
accesotros	12 (4.30%)	267 (95.70%)
privado	35 (12.54%)	244 (87.46%)
practodas	131 (46.95%)	148 (53.05%)
teoriatodas	122 (43.73%)	157 (56.27%)
horas2	162 (58.06%)	117 (41.93%)
considera23	79 (28.31%)	200 (71.69%)
tutoria5	9 (3.22%)	270 (96.78%)

4.1 Estimación del modelo logit clásico y Bayesiano

En primer lugar, se ha comparado la estimación Bayesiana no informativa de un modelo logit con la estimación frecuentista del modelo logit con el objeto de verificar que ambos modelos ofrecen resultados similares en términos de la estimación de parámetros. Posteriormente, se han comparado estos resultados con la estimación Bayesiana asimétrica del modelo logit y se comprueba que esta estimación Bayesiana asimétrica recoge el evidente efecto de asimetría que existe en los datos además de mejorar notablemente el ajuste de la probabilidad de aprobar. Se analizan también las evaluaciones de todos los modelos.

Para verificar la calidad del ajuste para el modelo logit estándar y los modelos Bayesianos, proponemos dos medidas diferentes: (1) el porcentaje de clasificación correcta en términos de aprobar la asignatura; (2) una medida estadística de ajuste como es el criterio de información de Akaike (AIC), o el criterio de información la desviación (DIC). Este último es particularmente útil en problemas de selección para modelos Bayesianos donde se usa simulación Markov Chain Monte Carlo (MCMC).

La distribución *a posteriori* para los modelos Bayesianos fue simulada en WinBUGS. Se realizaron un total de 100.000 iteraciones (después de un período de quemado “burn-in” de otras 100.000 simulaciones). Se ejecutaron cinco cadenas y se evaluó la convergencia para todos los parámetros con los distintos tests incluidos en el software WinBUGS (Convergence Diagnostics and Output Analysis).

Los resultados de estimación del modelo estándar logit desde el punto de vista frecuentista vienen dados en la Tabla 3. Como puede observarse en dicha Tabla, obtenemos 6 variables explicativas significativas al 1%: constante, contenidos, extras, privado, teoritodas y nonecesito. Obtenemos 3 variables explicativas significativas al 5%: matvirtual, humani y practodas. El test Chi Cuadrado muestra la significatividad global del modelo. Las variables significativas que se relacionan negativamente con la probabilidad de aprobar (frente a la probabilidad de no apro-

bar) son: que el estudiante considere que los contenidos de la asignatura le han resultado de mucha dificultad y que su opción de procedencia sea el Bachillerato de Humanidades. Sin embargo, la probabilidad relativa de aprobar la asignatura aumenta si: el alumno considera que el material disponible en el Campus Virtual es adecuado para el seguimiento de la asignatura, si el alumno ha asistido a clases extras, si el alumno proviene de un instituto de secundaria privado y, como era de esperar, si el alumno ha asistido a todas las clases de teoría y práctica.

La estimación Bayesiana estándar del modelo de regresión logit se resume en la Tabla 4. Podemos observar, como era de esperar (la información *a priori* es no informativa y el número de observaciones es alto) que las estimaciones de los parámetros son similares a las obtenidas en el modelo frecuentista estándar. Observamos que si consideráramos el 90% de credibilidad, entrarían más variables como relevantes, pero seguiremos considerando un nivel del 95%. De esta forma, hemos obtenido similares conclusiones en términos de evaluación de relevancia de las variables en la probabilidad de aprobar que en el modelo frecuentista.

Para nuestra base de datos, hemos obtenido un AIC de 269.166 para el método de estimación frecuentista, y un DIC de 270.135 para el modelo Bayesiano. Además, obtenemos prácticamente el mismo porcentaje de clasificación correcta en los dos modelos analizados. A la vista de las Tablas 3 y 4, podemos observar cómo para el modelo frecuentista la proporción ronda el 79.6%, mientras que para el método de estimación Bayesiano, dicho porcentaje se coloca en 79.2%. De nuevo, esto confirma que los dos modelos son similares también en términos de ajuste.

4.2 Estimación del modelo logit Bayesiano asimétrico

En la Tabla 5, se resumen los resultados de la estimación Bayesiana no informativa en la que se incluye una covariable que mide la posible asimetría del modelo. Como puede observarse en dicha Tabla, obtenemos prácticamente las mismas vari-

Tabla 3: Resultados del modelo Logit frecuentista (clásico)

Variables	Coeficientes	sd	p-valor
Constante	-1.6538	0.6216	0.008***
Contenidos	-1.0705	0.3894	0.006***
interés	0.6391	0.3602	0.076*
matvirtual	0.9438	0.3757	0.012**
prometeo	-0.9923	0.5280	0.060*
videos	0.4835	0.5280	0.355
extras	1.4538	0.3855	0.000***
tecnológico	0.8651	0.6113	0.157
humani	-1.3110	0.5257	0.013**
ciclofor	1.0203	0.7392	0.168
acesotros	-1.7076	1.2917	0.186
privado	1.4651	0.5289	0.006***
practodas	0.7903	0.4055	0.050**
teoriatodas	1.0701	0.3716	0.004***
horas2	-0.5967	0.3374	0.077*
considera23	-0.5538	0.4007	0.167
tutoria5	0.3244	0.7479	0.664

Vble dep.=Aprobar; $n = 279$; AIC=269.166; % Cl. correcta=79.6

***: 99% de nivel de confianza

**: 95% de nivel de confianza

*: 90% de nivel de confianza

Tabla 4: Resultados del modelo Logit Bayesiano estándar

Variables	Coeficientes	sd	Error MC	IC (95 %)
Constante	-1.8240	0.6538	0.0085	(-3.1310, -0.5869)***
Contenidos	-1.1710	0.3990	0.0030	(-1.9600, -0.3934)***
interés	0.7143	0.3845	0.0026	(-0.0339, 1.4700)*
matvirtual	1.0330	0.3720	0.0020	(0.3108, 1.7720)***
prometeo	-1.0802	0.5965	0.0078	(-2.2650, 0.0798)*
videos	0.5498	0.5983	0.0073	(-0.5929, 1.7570)
extras	1.5980	0.4057	0.0022	(0.8186, 2.4140)***
tecnológico	0.9482	0.6354	0.0030	(-0.2775, 2.2170)*
humani	-1.4640	0.5829	0.0023	(-2.6610, -0.3678)***
ciclofor	1.0790	0.7170	0.0031	(-0.3458, 2.4760)*
accesotros	-2.0040	1.1910	0.0042	(-4.5430, 0.1217)*
privado	1.6110	0.5445	0.0030	(0.5687, 2.7020)***
practodas	1.1670	0.4112	0.0028	(0.3706, 1.9840)***
teoriatodas	0.8696	0.4151	0.0027	(0.0547, 1.6900)**
horas2	-0.6577	0.3616	0.0022	(-1.3660, 0.0462)*
considera23	-0.6164	0.4028	0.0018	(-1.4170, 0.1651)*
tutoria5	0.4058	0.9676	0.0037	(-1.4570, 2.3501)

Vble dep.=Aprobar; $n = 279$; DIC=270.135; % Cl. correcta=79.2

***: 99% de nivel de credibilidad

**: 95% de nivel de credibilidad

*: 90% de nivel de credibilidad

ables relevantes con respecto a los dos métodos de estimación anteriores. Podemos observar cómo aparece una nueva variable relevante, como es “accesotros” (la probabilidad relativa de aprobar disminuye si el estudiante proviene de otras opciones de acceso a la universidad, frente a la de Ciencias Sociales) y perdemos la relevancia al 95% de la variable “teoriatodas”. Obviamente, el signo de los coeficientes se mantiene, si bien observamos cómo los valores de los mismos difieren notablemente con respecto a los obtenidos en los modelos de estimación frecuentista y Bayesiano estándar. Este hecho se acentúa aún más en la estimación de la constante, por lo que observamos que en los dos primeros modelos, la constante estimada recoge parte del efecto asimetría que ha salido a la luz en el tercer modelo. Además, observamos también cómo la variable δ que mide la asimetría de los datos es estadísticamente relevante y positiva, indicando que es necesario tener en consideración la asimetría existente en los datos a la hora de especificar el modelo de estimación logit. Así, este coeficiente δ está ajustando la probabilidad estimada de aprobar, es decir, aumentando la probabilidad de aprobar de cada estudiante, captando de esta forma la menor cantidad de información con $Y_i = 1$. Este hecho, con total seguridad, mejorará el ajuste en la estimación de la probabilidad de aprobar.

El criterio DIC es igual a 78.187 para este modelo asimétrico, es decir, notablemente menor que el obtenido para el modelo simétrico (270.135). Esta reducción tan importante en esta medida indica un incremento significativo en el nivel de ajuste del modelo. Además, se observa igualmente una mejora en la capacidad de clasificación de los datos, ya que el nivel de porcentaje de ajuste de los aprobados, es perfecto (100%). Obviamente, el incremento en la probabilidad del ajuste de los casos $y_i = 1$ inducido por el coeficiente de asimetría δ positivo explica estos resultados tan buenos. El modelo frecuentista y el bayesiano asimétrico estándar fallaban en el ajuste, sobre todo, de $y_i = 1$.

Tabla 5: Resultados del modelo Logit Bayesiano asimétrico

Variables	Coefficientes	sd	Error MC	IC (95 %)
Constante	-23.8000	4.9140	0.1869	(-34.4700, -15.1500)***
Contenidos	-8.2520	3.0610	0.0738	(-14.8700, -2.8530)***
interés	3.5830	2.4110	0.0541	(-0.9988, 8.5560)*
matvirtual	5.4370	2.3420	0.0517	(1.1710, 10.4300)**
prometeo	-5.6580	3.6620	0.1066	(-13.2400, 1.1770)*
videos	3.2750	3.5320	0.0962	(-3.6740, 10.3900)
extras	8.8060	2.8200	0.0604	(3.6320, 14.6900)***
tecnológico	3.8590	3.5770	0.0511	(-3.088, 11.07)
humani	-7.8380	3.9460	0.0546	(-16.2800, -0.8753)**
ciclofor	3.3620	4.0950	0.0456	(-5.0450, 11.2000)
accesotros	-11.3700	5.4720	0.0557	(-22.6900, -1.2810)**
privado	8.1040	3.2090	0.0608	(2.181, 14.77)***
practodas	6.7970	2.9700	0.0731	(1.4200, 13.05)**
teoriatodas	4.7100	2.7790	0.0613	(-0.4682, 10.61)*
horas2	-3.7260	2.3670	0.0440	(-8.6890, 0.6575)*
considera23	-3.8250	2.5190	0.03829	(-9.0800, 0.8222)*
tutoria5	0.5993	5.0330	0.0506	(-9.7970, 10.0800)
delta	19.0400	3.8170	0.1353	(12.0500, 27.1600)***

Vble dep.=Aprobar; $n = 279$; DIC=78.187; % Cl. correcta=100

***: 99% de nivel de credibilidad

**: 95% de nivel de credibilidad

*: 90% de nivel de credibilidad

5 CONCLUSIONES

Los resultados obtenidos con los tres modelos utilizados son muy similares, destacando que en ellos las variables que determinan la probabilidad relativa de aprobar la asignatura son: la asistencia con regularidad a las clases de teoría y práctica, el material disponible en el Campus Virtual, si el alumno ha asistido a clases extras, si el alumno proviene de un instituto de secundaria privado.

Por otra parte, afectan negativamente a la probabilidad de aprobar, la dificultad que los estudiantes encuentran en los contenidos, la procedencia del bachillerato de humanidades, y en el modelo bayesiano asimétrico, se incorpora además la procedencia a través de otro tipo de acceso.

Por su parte, el modelo asimétrico Bayesiano capta perfectamente la asimetría existente en los datos, ponderando el ajuste de la probabilidad para ajustar también los casos con menor frecuencia.

Llaman la atención cómo afectan positivamente a la probabilidad de aprobar las variables centro privado y las clases extras. En cambio, con respecto a las variables horas dedicadas diariamente al estudio de la asignatura y las horas de estudio que el estudiante considera que semanalmente debería dedicar a la materia, sólo se muestran como significativas con un nivel de credibilidad al 90% en los modelos bayesianos. Entendemos que este resultado es consecuencia de cómo agrupamos los datos de la muestra.

Esta es la línea de trabajo con la que continuamos. Nos estamos centrando en las horas de trabajo personal del alumno, que definen junto con la dedicación del profesorado los créditos de la materia. En este sentido los nuevos resultados apuntan que los estudiantes dedican semanalmente menos del tiempo medio definido en la asignatura y que esta variable sí va afectar positivamente a la probabilidad de aprobar.

Concluyendo, si todos los factores que parecen incidir en el éxito o fracaso de

un estudiante de ME de GADE pudieran incorporarse en el modelo, entonces parece razonable que a la luz de los resultados pudieramos establecer pautas de actuación en el futuro que permitan corregir determinados comportamientos y que puedan llevar a un mayor éxito en dicha asignatura.

6 REFERENCIAS BIBLIOGRÁFICAS

- ALBERT, J.H. Y CHIB, S. (1993). “Bayesian Analysis of Binary and Polychotomous Response Data”. *Journal of the American Statistical Association*, 88 (422), 669–679.
- ALBERT, J.H. & CHIB, S. (1995). “Bayesian residual analysis for binary response regression models. *Biometrika* 82, 747–769.
- BASU, S. Y MUKHOPADHAYAY, S. (2000). “Binary response regression with normal scale mixture links”. In: Dey, D.K., Ghosh, S.K. & Mallick, B.K. eds. *Generalized Linear Models: A Bayesian Perspective*. New York: Marcel Dekker.
- BAZÁN, J.L., BRANCO, M.D. Y BOLFARINEZ, H. (2006). “A Skew Item Response Model”. *Bayesian Analysis*, 1 (4), 861–892.
- BERMÚDEZ, LL., PÉREZ, J.M., AYUSO, M., GÓMEZ, E. Y VÁZQUEZ, F.J. (2008). “A Bayesian dichotomous model with asymmetric link for fraud in insurance”. *Insurance: Mathematics and Economics* 42, 779–786.
- CARLIN, B.P. Y POLSON, N.G. (1992). “Monte Carlo Bayesian methods for discrete regression models and categorical time series”. *Bayesian Statistics* 4, 577–86.

- CHEN, M.H., DEY, D.K. Y SHAO, Q.M. (1999). “A New Skewed Link Model for Dichotomous Quantal Response Data”. *Journal of the American Statistical Association* 94, 1172–1186.
- CHEN, W.S., BAKSHI, B.R., GOEL, P.K. Y UNGARALA, S. (2004). “Bayesian estimation of unconstrained nonlinear dynamic systems via sequential Monte Carlo sampling”. *Industrial & Engineering Chemistry Research* 43 (14), 4012–4025.
- DÁVILA, N., GARCÍA, M.D. Y GÓMEZ, E. (2011). “Una reflexión sobre la evaluación en Matemáticas Empresariales”. *XIX Jornadas ASEPUMA–VII Encuentro Internacional Anales de ASEPUMA n° 19*: 403.
- FEDRIANI, E.M. Y ROMANO, I. (2011). “Causas del rendimiento académico en asignaturas cuantitativas de la Facultad de Ciencias Empresariales”. *XIX Jornadas ASEPUMA–VII Encuentro Internacional Anales de ASEPUMA n° 19*: 0603.
- FLORIDO, C., JIMÉNEZ, J. Y SANTANA, I. (2011). “Obstáculos en el camino hacia Bolonia: efectos de la implantación del Espacio Europeo de Educación Superior (EEES) sobre los resultados académicos”. *Revista de Educación*, 354, 629–653.
- GILKS, W.R., RICHARDSON S. Y SPIEGELHALTER D.J. (1995). “Introducing markov chain monte carlo”. In: Gilks, W.R., Richardson S., Spiegelhalter D.J., eds. *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.
- LUNN D.J., THOMAS A., BEST N. Y SPIEGELHALTER D. (2000). “WinBUGS: a Bayesian modelling framework: concepts, structure, and extensibility”. *Statistics and Computing*, 10, 325–37.

- NAGLER, J. (1994). “Scobit: An Alternative Estimator to Logit and Probit”. American Journal of Political Science, 38, 1, 230–255.
- PRENTICE, R. (1976). “A Generalization of the Probit and Logit Methods for Dose–Response Curves”. Biometrics, 32, 4, 761–768.
- RÚA, A., REDONDO R., MARTÍNEZ DE IBARRETA, C., FABRA M.E., MARTÍN, M.J. Y NÚÑEZ, A. (2010). “Factores del rendimiento académico en las asignaturas cuantitativas de Administración y Dirección de empresas”. XVIII Jornadas ASEPUMA–VI Encuentro Internacional Anales de ASEPUMA n° 18: 105.
- SÁEZ, J.A. OLMO, M.J., PÉREZ, J.M., NEGRÍN, M.A., ARCOS, A. Y DÍAZ, J. (2010). “Bayesian Analysis of Nosocomial Infection Risk and Length of Stay in a Department of General and Digestive Surgery”. Value in Health, 13, 4, 431–439.
- STUKEL, T. (1988). “Generalized logistic model”. Journal of the American Statistical Association, 83, 426–431.