

APLICACIÓN DE LOS MODELOS PROBABILÍSTICOS PARA ANALIZAR LA EFICIENCIA EN LAS UNIVERSIDADES

Angélica Beatriz Contreras Cueva

Enrique Cuevas Rodríguez

Universidad de Guadalajara

Catalina Bolancé Losilla

Universidad de Barcelona

ABSTRACT

End efficiency is defined as the student's ability to obtain a university degree within the expected length of time provided for, in the Studies Program. Based on the latter definition we propose a model to bring levels of end efficiency close together. The above mentioned model considers three types of students: the one who drops out of school, the one who graduates within the expected period of time set forth in the Studies Program and the one that finishes his/her studies way past said period of time. The model establishes that end efficiency depends on the number of semesters for which the student registers at the University and the likelihood that he/she may graduate. In order to make an approximate calculation we bring into play four cohorts of students which belong to the Economics and Administration Science Studies Center at the University of Guadalajara. The Logit model allows us to calculate, for each student, the likelihood to be persistent and achieve graduation within the maximum period of time established. The Poisson model allows us to calculate the expected number of semesters that the student will stay at the University. To end with, the estimation model allows us to predict an end efficiency figure for new students.

Palabras Clave: Eficiencia en la educación, modelos probabilísticos, predicción, economía de la educación

1. INTRODUCCIÓN

La eficiencia es la capacidad que un alumno tiene de concluir una titulación universitaria, la falta de eficiencia puede ser el resultado, entre otros aspectos, del abandono de los estudios y el retraso en la consecución del título, ya que esto supone para el sistema educativo importantes pérdidas de recursos, por ello, el objetivo de este trabajo, por una parte, es proponer un modelo para medir la eficiencia con el que se puede predecir en que medida un alumno podría resultar ineficiente para el sistema educativo, y por otra, proporcionar una herramienta que puede ayudar a asesorar al alumno en la elección de una titulación universitaria.

El modelo propuesto se sustenta en que la eficiencia depende de una serie de probabilidades tales como: la probabilidad de abandonar los estudios, la de titularse y la de retrasar excesivamente la consecución del título universitario. Y también, del número esperado de semestres que el alumno permanece matriculado. Finalmente, se puede predecir el nivel de eficiencia para cada titulación dadas las características conocidas previamente del futuro alumno.

En este trabajo primero se describe el modelo de eficiencia y la forma de aproximarlos. Posteriormente, se describen la base de datos utilizada en la estimación del modelo y algunos resultados obtenidos. Finalmente, en un último apartado se resumen los principales resultados y conclusiones.

2. EL MODELO DE EFICIENCIA UNIVERSITARIA

La eficiencia en la obtención del título universitario está ligada, por un lado, a la probabilidad de que el alumno abandone los estudios y, por otro, a la que no abandone, este segundo término involucra dos aspectos fundamentales el que el estudiante finalice sus estudios satisfactoriamente y el tiempo que este tarda en conseguirlo. Tomando en cuenta estos aspectos se obtiene el siguiente indicador de eficiencia en la consecución del título universitario.

$$E_i = \frac{\left(p_i \sum_{j=0}^n j p_{ij}^B \right) + (1 - p_i) \left[p_i^T a_T + (1 - p_i^T) \sum_{k=1}^m k p_{ik}^T \right]}{a_T}, \quad (1)$$

donde p_i es la probabilidad de que el alumno i abandone, p_i^T es la probabilidad de titularse, a_T es el número de semestres considerados como adecuados por el plan de estudios para titularse, p_{ij}^B es la probabilidad de que el alumno i abandone en el semestre j y p_{ik}^T es la probabilidad de que el alumno i retrase su titulación en k semestres. Cuando se hace referencia al abandono en el semestre de ingreso, sin cubrir ningún crédito, j vale 0, en el primer semestre teniendo algún crédito aprobado j vale 1, en el segundo 2 y así sucesivamente.

Dada la expresión (1), se observa que para el cálculo de la eficiencia E_i es necesario disponer del valor de las probabilidades que intervienen en el modelo. A continuación, se describe como se estiman dichas probabilidades.

Las probabilidades p_i y p_i^T se obtienen mediante el ajuste de dos modelos logit, el primero relaciona la probabilidad de que el alumno con determinadas características abandone o no abandone los estudios. El segundo modelo logit relaciona la probabilidad de titularse dado que el alumno no abandonó los estudios. Definidos los modelos utilizados, los valores teóricos de las probabilidades son:

$$p_i = \frac{\exp(X_i\beta)}{1 + \exp(X_i\beta)} \text{ y } p_i^T = \frac{\exp(X_i\tau)}{1 + \exp(X_i\tau)}, \quad (2)$$

donde X_i es un vector fila con el valor de las variables explicativas para el alumno i y β y τ son vectores columna de parámetros a estimar.

Las probabilidades p_{ij}^B y p_{ik}^T se estiman a partir del ajuste de modelos de Poisson. Por un lado se relaciona la variable número de años que el alumno con determinadas características tarda en abandonar y, por otro, bajo la condición de que el alumno ni ha abandonado ni se ha titulado en un periodo de 5 años, se ajusta la variable número de semestres que el alumno tarda en titularse. Los valores de las probabilidades se obtienen como:

$$p_{ij}^B = \frac{e^{-\exp(X_i\alpha)} [\exp(X_i\alpha)]^j}{j!} \text{ y } p_{ij}^T = \frac{e^{-\exp(X_i\theta)} [\exp(X_i\theta)]^j}{j!}, \quad (3)$$

donde α y θ son vectores columna de parámetros a estimar. En todos los modelos, las variables explicativas utilizadas son las mismas, todas ellas se describen posteriormente.

Teniendo en cuenta la expresión (1), el valor mínimo para E_i es igual a 0 y se da cuando las probabilidades p_i y p_{i0}^B son iguales a 1 y el resto son cero. El valor máximo de E_i es superior a la unidad y depende del máximo de k . En la práctica, para los alumnos que se abandonan los valores de E_i tienden a ser pequeños, por el contrario, para los alumnos que se titulan en a_T semestres el valor de la eficiencia está próximo a 1. Finalmente, la eficiencia para los alumnos que tardan en titularse más de a_T semestres se reparte a lo largo de la distribución de E_i , esto sucede porque en estos casos las probabilidades de abandonar pueden ser elevadas.

Cabe mencionar que existe un problema de censura para aquellos alumnos que, después de haber finalizado el periodo de recogida de datos para la investigación, continúan matriculados, es decir en caso de ser positivo m no se conoce en su totalidad. Por lo tanto, se sabe que: $m = m^c + c$, donde, m^c es el valor conocido. El valor de m es desconocido y equivale al número de semestres que

transcurren hasta que finalmente el alumno se titula o abandona. La estimación de m se realiza bajo el supuesto de que el alumno que continúa matriculado conseguirá titularse una vez concluidos los créditos necesarios para obtener el título. El número de semestres que restan hasta dicho momento c se estima teniendo en cuenta, por un lado, los créditos que le faltan por cursar en el último semestre de recogida de la información y, por otro, el promedio de créditos que aprueba por semestre.

Tabla 1: Variables en la base de datos.

VARIABLE	DESCRIPCIÓN
TIPO	Toma valores: 1=abandona, 2=titula y 3=continua matriculado
BAJA	Toma valores: 1=abandona y 0 en caso contrario
ADM	Toma valores: 1=Titulado en Administración y 0 en caso contrario
AFS	Toma valores: 1=Titulado en Administración Financiera y Sistemas y 0 en caso contrario
CPU	Toma valores: 1=Titulado en Contaduría Pública y 0 en caso contrario
ECO	Toma valores: 1=Titulado en Economía y 0 en caso contrario
MER	Toma valores: 1=Titulado en Mercadotecnia y 0 en caso contrario
NIN	Toma valores: 1=Titulado en Negocios Internacionales y 0 en caso contrario
RHU	Toma valores: 1=Titulado en Recursos Humanos y 0 en caso contrario
SIN	Toma valores: 1=Titulado en Sistemas de Información y 0 en caso contrario
TUR	Toma valores: 1=Titulado en Turismo y 0 en caso contrario
HOMBRE	Toma valores: 1=Hombre y 0 en caso contrario
PUBLICA	Toma valores: 1=procedencia escuela pública y 0 en caso contrario
EDAD	Edad del estudiante en el momento de matricularse
CREDITOS	Número de créditos aprobados hasta el momento de recogida de datos
SEMESTRES	Número de semestres en los que matricula alguna asignatura hasta el momento de recogida de datos
PAA	Nota de la prueba de acceso a la universidad (toma valores entre 60 y 100)
PROM_PRE	Nota media del bachillerato (toma valores entre 60 y 100)
PER_MAT	Periodo de matrícula: 1=enero y 0=agosto

Es importante tener en cuenta que el cálculo del indicador de eficiencia puede ayudar a asesorar a futuros alumnos en la elección de estudios. Es decir, el modelo se plantea con objetivos predictivos, aunque en este trabajo se presentan los resultados para una muestra de alumnos que ya se han titulado o abandonaron los estudios. A partir de estos datos estimaremos los vectores de parámetros que definen los modelos logit y de Poisson, para finalmente predecir las probabilidades de

abandono y de titularse en cinco años. A continuación, en la siguiente sección se describe la base de datos utilizada.

3. LA BASE DE DATOS

Para estimar los valores de la eficiencia terminal propuesta en la expresión (1) se utiliza una base de datos que contiene información referida a cuatro cohortes de estudiantes del Centro Universitario de Ciencias Económico Administrativas de la Universidad de Guadalajara (México). Los alumnos analizados se matricularon en los cursos de febrero y agosto de 1998 y 1999, respectivamente, y suman un total de 6480. En la tabla 1 se describen las variables que forman la base de datos analizada.

4. RESULTADOS OBTENIDOS

A continuación, se presentan los resultados de los modelos logit y de Poisson, que se estiman para calcular las probabilidades que intervienen en el indicador de eficiencia terminal E_i , expresado en (1). Todos los resultados se obtienen con el programa estadístico SAS. Los resultados del modelo logit estimado para calcular la probabilidad de abandono se muestran en el cuadro 1.

Los resultados del cuadro 1 permiten analizar el efecto de las variables explicativas en la probabilidad de abandono. Analizando aquellos parámetros que resultan significativamente distintos de cero, en general, se observa que las titulaciones con mayor probabilidad de abandono son Economía y Turismo. Asimismo, cuanto mayor en edad es el alumno en el momento de matricularse mayor es la probabilidad de darse de abandonar. Con respecto al bachillerato de procedencia, la probabilidad de abandono disminuye si el alumno procede de escuela pública. Además, se observa que los hombres tienden a abandonar más los estudios que las mujeres. Por otra parte, la nota de acceso y el promedio del bachillerato se observa que en cuanto menores sean estos inciden en la probabilidad de abandonar los estudios. Finalmente, los alumnos matriculados en el mes de febrero abandonan más que aquellos que ingresan en agosto.

Cuadro 1: Estimación del modelo logit para el ajuste de la probabilidad de abandonar p_i .

Número de observaciones	6480	Nombre de la distribución:	Logistic
Verosimilitud Log	3704.215991	Perfil de respuesta: abandona	2102 no abandona 4378
Parámetro	Estimador	Chi-cuadrada	Pr > ChiSq
Intercept	1.1697	4.71	0.0300
Administración Financiera y Sistemas	-0.0233	0.02	0.8880
Contaduría Pública	-0.4397	20.29	0.0001
Economía	1.4214	97.81	0.0001
Mercadotecnia	0.1764	1.90	0.1684
Negocios Internacionales	-0.0456	0.12	0.7282
Recursos Humanos	0.1252	0.55	0.4585
Sistemas de Información	-0.0144	0.01	0.9220
Turismo	0.5779	25.60	0.0001
Pública	-0.2573	14.68	0.0001
Edad	0.1179	150.06	0.0001
Hombre	0.1917	9.49	0.0021
PAA	-0.0047	2.72	0.0992
Prom_Pre	-0.0446	76.58	0.0001
Permat	-0.0967	2.61	0.1063

Los resultados de la estimación del modelo logit que permite calcular la probabilidad p_i^T de que el alumno se haya titulado en el periodo a_T previsto por el plan de estudios se presentan en el cuadro 2, estos resultados se obtienen con la muestra de alumnos que no han abandonado durante el periodo de análisis. En el cuadro 2 se muestran los resultados obtenidos tomando como categoría de referencia el hecho de continuar matriculado una vez transcurrido el periodo de estudio. Por tanto, la interpretación de los parámetros estimados debe realizarse en función de la probabilidad de continuar.

La probabilidad de titularse en a_T semestres se obtiene como el contrario de la estimada por el modelo logit presentado en el cuadro 2. Analizando los signos de los parámetros estimados se observa como los individuos que no se titulan en el tiempo previsto por el plan de estudios tiene características similares a los que abandonan los estudios. Sin embargo, hacemos hincapié en los siguientes aspectos, los estudiantes de Economía y los de Turismo, los hombres, los que provienen de bachilleratos públicos y los que se matriculan en el mes de febrero consumen más tiempo en conseguir el título. Sobresale también, que los alumnos matriculados en Contaduría Pública, Sistemas de Información, Negocios Internacionales, Mercadotecnia y aquellos con altos promedios en bachillerato finalizan sus estudios en el tiempo considerado como adecuado.

Cuadro 2: Estimación del modelo logit para el ajuste de la probabilidad P_i^T de titularse en a_T semestres.

Número de observaciones	4378	Nombre de la distribución	Logistic
Verosimilitud Log	-2112.950707	Perfil de respuesta: continua	1077 titula 3301
Parámetro	Estimador	Chi-cuadrada	Pr > ChiSq
Intercept	3.9368	25.54	0.0001
Administración Financiera y Sistemas	-0.4973	6.10	0.0135
Contaduría Pública	-1.7146	198.47	0.0001
Economía	0.5400	6.19	0.0129
Mercadotecnia	-0.5039	10.65	0.0011
Negocios Internacionales	-0.8480	28.87	0.0001
Recursos Humanos	0.0228	0.01	0.9085
Sistemas de Información	-1.4979	51.65	0.0001
Turismo	0.3230	5.22	0.0223
Pública	0.3385	11.56	0.0007
Edad	0.0212	1.93	0.1646
Hombre	0.3553	17.82	0.0001
PAA	0.0075	3.40	0.0652
Prom_Pre	-0.0653	84.03	0.0001
Permat	0.2054	6.25	0.0124

Los resultados de los modelos de Poisson estimados para el cálculo P_{ij}^B y P_{ik}^T se presentan, respectivamente, en los cuadros 3 y 4. En el primero, la variable dependiente es el número de semestre en el que abandona y, en el segundo modelo de Poisson, la variable dependiente es el número de semestres m que transcurren desde a_T hasta que se titula.

La estimación del cuadro 3 se realiza con aquellos alumnos que abandonan y, la del cuadro 4, se obtiene con aquellos alumnos que continúan matriculados tras finalizar el periodo de recogida de datos.

Cuadro 3: Estimación del modelo de Poisson para el ajuste de la probabilidad P_{ij}^B .

Distribución Poisson	Función de vínculo		Log
Variable dependiente Semestres	Total de observaciones leídas		2102
Parámetro	Estimador	Chi-cuadrada	Pr > ChiSq
Intercept	4.2410	302.56	0.0001
Administración Financiera y Sistemas	-0.0477	0.45	0.5044
Contaduría Pública	-0.3420	52.92	0.0001
Economía	-0.7270	144.71	0.0001
Mercadotecnia	0.0412	0.53	0.4677
Negocios Internacionales	-0.0163	0.07	0.7878
Recursos Humanos	-0.1016	1.73	0.1882
Sistemas de Información	-0.0183	0.08	0.7830
Turismo	-0.3229	37.74	0.0001
Pública	0.1010	10.07	0.0015
Edad	-0.0337	59.65	0.0001
Hombre	0.1987	44.76	0.0001
PAA	-0.0030	6.05	0.0139
Prom_Pre	-0.0277	135.35	0.0001
Permat	0.0013	0.00	0.9617

Con relación a los estudiantes que abandonan, según el cuadro 3, podemos señalar que los estudiantes matriculados en las titulaciones de Contaduría Pública, Economía y Turismo abandonan los estudios los primeros semestres. Los estudiantes menores, en cuanto a la edad, que abandonan los hacen iniciando la licenciatura. Los que provienen de bachilleratos públicos y los hombres que abandonan, retardan más el abandono. Finalmente los estudiantes que tienen menores notas del bachillerato abandonan los primeros semestres. Por otra parte, con respecto a los estudiantes que se titularon a tiempo, según el cuadro 4, sobresale que los que se matriculan en el calendario de febrero consumen mayor tiempo que los que lo hacen en el calendario de agosto. En cuanto a la edad, los estudiantes mayores consumen más tiempo en titularse.

Una vez estimadas las probabilidades que intervienen en la expresión (1), se está en disposición de obtener un valor de eficiencia terminal E_i para cualquier alumno i en este caso se calculan las eficiencia para los alumnos que forman parte de la base de datos con la que se ha realizado la estimación. Aunque, el modelo podría aplicarse a cualquier alumno que quisiera matricularse en el centro objeto de estudio.

Cuadro 4: Estimación del modelo de Poisson para el ajuste de la probabilidad P_{ik}^T .

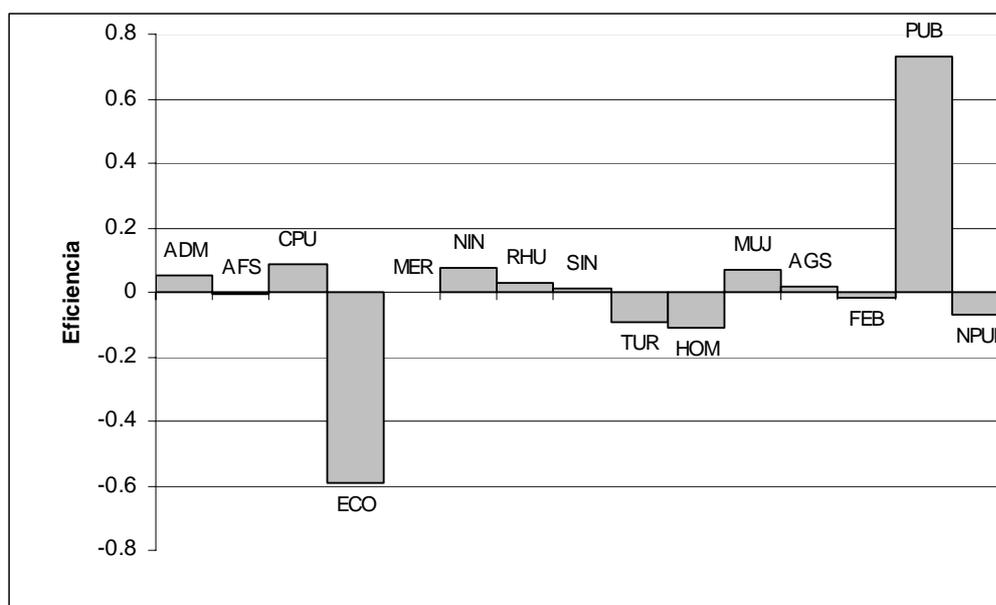
Distribución Poisson		Función de vínculo	
Variable dependiente	Semestres	Total de observaciones leídas	Log 597
Parámetro	Estimador	Chi-cuadrada	Pr > ChiSq
Intercept	1.3388	8.85	0.0029
Administración Financiera y Sistemas	-0.0085	0.011	0.9409
Contaduría Pública	-0.0835	1.19	0.2758
Economía	-0.1432	1.87	0.1715
Mercadotecnia	-0.0214	0.05	0.8262
Negocios Internacionales	0.0385	0.16	0.6864
Recursos Humanos	-0.1311	1.30	0.2550
Sistemas de Información	-0.0688	0.22	0.6425
Turismo	-0.0482	0.35	0.5513
Pública	-0.0201	0.12	0.7344
Edad	0.0229	7.75	0.0064
Hombre	0.0352	0.47	0.4917
PAA	-0.0020	0.68	0.4111
Prom_Pre	-0.0071	2.68	0.1014
Permat	0.2075	16.50	0.0001

A continuación, en el cuadro 5 se muestran algunos estadísticos descriptivos del indicador de eficiencia, para el total de la muestra y, también, en función de si el estudiante abandonó, se tituló en el periodo previsto por el plan de estudios o continuaba matriculado al finalizar el periodo de recogida de datos. Observar como, en promedio, para los alumnos que han abandonado la eficiencia es menor. Sin embargo, entre los alumnos que se titulan y los que continúan no se observan excesivas diferencias en promedio.

Cuadro 5: Estadísticos descriptivos de la eficiencia.

Total				
Número de observaciones	Media	Mediana	Máximo	Mínimo
6480	0.8122	0.8365	1.2738	0.1045
Abandonan				
2102	0.7887	0.8191	1.2738	0.1045
Titulados				
3781	0.8246	0.8416	1.2228	0.2109
Continuaron				
597	0.8166	0.8349	1.2424	0.2509

Gráfico 1: Eficiencia para las variables



Finalmente del gráfico 1 podemos señalar que las titulaciones con mayor eficiencia fueron Contaduría Pública, Negocios Internacionales y Administración, contrario a ello, las menos eficientes fueron las de Economía y Turismo, las mujeres presentaron más eficiencia en los estudios que los hombres, los estudiantes matriculados en agosto y procedentes de bachilleratos públicos resultaron ser más eficientes.

5. CONCLUSIONES

De las principales conclusiones se resalta que el modelo permite identificar el nivel de eficiencia de las titulaciones de las cohortes analizadas, esto constituye un importante punto de partida para sustentar las bases para establecer criterios de mejora para la institución en general y para las titulaciones en particular. Finalmente, La herramienta propuesta en este trabajo puede ser de utilidad en el asesoramiento del alumno antes de matricularse en una titulación. Puede establecerse que si un determinado alumno no alcanza un cierto grado de eficiencia en la titulación que quiere estudiar, es objetivo de asesorarle en otra dirección, aunque, finalmente, sea el propio alumno el que tome la decisión.

6. REFERENCIAS BIBLIOGRÁFICAS

Alemany, R. (1993), *Modelització de la Durada dels Estudis Universitaris: Una Aplicació a la Universitat de Barcelona*, Tesis Doctoral, Universidad de Barcelona, España.

Contreras, A.B. y Bolancé, C. (2005), “El coste temporal del abandono de los estudios en el centro universitario de Ciencias Económico Administrativas de la Universidad de Guadalajara”, Ed. Umbral Digital Publicaciones, Universidad de Guadalajara, México.

Contreras, A.B. et al. (2006).”Eficiencia terminal en las universidades”, *Expresión Económica, Revista de Análisis; Centro de Investigaciones Sociales y Económicas*, Universidad de Guadalajara, 15. pp. 41-53.

Contreras, A. B. (2006), “Remuneraciones a la educación, desigualdad educativa y abandono escolar”, *Compilación*, pp. 1-20.