

UN MODELO DE SUPERVIVENCIA PARA DATOS DISCRETOS EN LA IDENTIFICACIÓN DE FACTORES QUE AFECTAN EL RENDIMIENTO ACADÉMICO UNIVERSITARIO

Juan Carlos Salazar Uribe, Carlos Mario Lopera, Mario Cesar Jaramillo
Elorza

Escuela de Estadística, Universidad Nacional de Colombia, Sede
Medellín

2010

Resumen

El fenómeno de la pérdida de calidad de estudiante por bajo rendimiento académico es un problema serio que afecta las principales universidades del país. El bajo rendimiento académico también está relacionado de manera directa con la deserción estudiantil. Por lo tanto es de fundamental importancia identificar factores que pueden estar asociados con la pérdida de calidad de estudiante debida a bajo rendimiento académico. Con base en información sobre rendimiento académico recopilada en la Universidad Nacional de Colombia, Sede Medellín se implementó un modelo de análisis de supervivencia para tiempo discreto basado en la transformación logit donde la respuesta considerada fue el tiempo que transcurre hasta que un individuo pierde la calidad de estudiante por primera vez. Este modelo permitió identificar y cuantificar el efecto de variables de tipo socio-demográfico relacionadas con la pérdida de calidad de estudiante.

Palabras Clave: Análisis de supervivencia, razón de odds, rendimiento académico.

Abstract

Losing academic privileges due to poor academic performance is a serious problem affecting Colombian top universities. The poor academic performance is also related directly to student dropout. Therefore it is of fundamental importance to identify factors that may be associated with the losing of academic privileges due to poor academic performance. Based on academic performance information collected at the National University of Colombia at Medellín we built a survival

analysis model based on a logit transformation for discrete time where the endpoint considered was the time that elapses before an individual loses academic privileges for the first time. This model allowed us to identify and quantify the effect that socio-demographic variables could have on the lost of academic privileges.

Keywords: Survival analysis, odds ratio, academic performance.

1. INTRODUCCIÓN

Los modelos de análisis de supervivencia, por ejemplo el modelo semi-paramétrico de Cox, se usan con frecuencia en muchas áreas de investigación que incluyen epidemiología, ciencias médicas y sociales. Estos modelos tienen la ventaja de que tienen en cuenta el tiempo que transcurre hasta la ocurrencia de un evento de interés y además permiten incorporar información relacionada con censura a derecha la cual se presenta cuando el evento no ocurre durante el tiempo de observación. Se han reportado múltiples estudios en la literatura donde se identifican factores asociados con el tiempo de ocurrencia de un evento de interés. Estos modelos asumen que el tiempo que transcurre hasta el evento es una variable aleatoria de tipo continuo. Sin embargo, para tratar de identificar factores posiblemente relacionados con el tiempo que transcurre hasta que un individuo pierde la calidad de estudiante por primera vez, este supuesto puede no ajustarse a la realidad ya que durante su paso por la Universidad un estudiante experimenta recesos debidos principalmente a períodos de vacaciones .

En la literatura se pueden encontrar otros trabajos relacionados con la implementación de modelos y métodos estadísticos para analizar el rendimiento académico. Por ejemplo, Correa y Lopera (2008) discuten una técnica basada en tablas de vida para modelar la mortalidad estudiantil universitaria. Tovar y Salazar (2007) ilustran un modelo lineal mixto con estructura markoviana de primer orden usando datos sobre rendimiento académico en la Universidad Nacional de Colombia, Sede Medellín. Por otra parte Castaño et al (2006) aplicaron modelos de duración para analizar los factores asociados a la deserción y graduación estudiantil universitaria.

En este trabajo se utilizó un modelo de supervivencia para datos discretos para identificar algunos de los factores que afectan la pérdida por primera vez de la calidad de estudiante en la Universidad Nacional

de Colombia, Sede Medellín. El modelo discreto se tomó por la forma en que se dan los periodos académicos en la universidad y se espera que es te trabajo y trabajos futuros en la misma dirección sirvan como base para evaluar políticas académicas y así soporte de una mejor forma la toma de decisiones.

La sección 2 describe brevemente la base de datos utilizada para la aplicación del modelo de supervivencia para datos discretos. El modelo utilizado está explicado en la sección 3. Los resultados obtenidos son presentados en sección 4. Finalmente, en la sección 5 se dan algunas conclusiones.

2. BASE DE DATOS

Se va a trabajar con una base de datos recolectada en la Universidad Nacional durante los años 2003 a 2008, la cual contiene información semestre a semestre de variables sociodemográficas y de rendimiento académico de 12786 estudiantes. La información de esta base de datos incluye además aquella relacionada con el causal de pérdida de calidad de estudiante, que puede deberse básicamente a bajo rendimiento académico, o por haber culminado exitosamente el ciclo de estudios de sus respectivas carreras.

Esta información permitió definir para cada uno de estos estudiantes un tiempo transcurrido en semestres hasta que él o ella pierden la calidad de estudiante por primera vez. La información de este tiempo para estudiantes que culminaron su ciclo de estudios exitosamente se trato como censura a derecha.

3. MODELO

En el análisis de datos sobre promedios académicos no parece razonable asumir que el tiempo que transcurre hasta que una persona pierde la calidad de estudiante por primera vez sea continuo ya que durante el tiempo de permanencia en la Universidad hay interrupciones ocasionadas, por ejemplo, por los períodos de vacaciones. Por esta razón es aconsejable tratar dicho tiempo en una escala discreta donde

los semestres representan puntos discretos en el tiempo a los cuales el evento de interés ocurre de una manera regular.

Este tiempo discreto no debe modelarse con un modelo de Cox sino más bien con un modelo de regresión logística que permita estimar la probabilidad condicional de que una persona, por ejemplo la i -ésima, experimente el evento de interés en un tiempo t dado que el evento no ha ocurrido antes de t . Esta probabilidad usualmente se denota P_{it} . Este modelo establece que esta probabilidad condicional se relaciona con las covariables por medio de la siguiente ecuación (Allison 1995):

$$\log\left(\frac{P_{it}}{1-P_{it}}\right) = \alpha_i + \sum_{j=1}^p \beta_j x_{ij}, \quad i=1, \dots, n \text{ y } j=1, \dots, p \quad (1)$$

En esta ecuación el intercepto α_i , que se refiere al log-odds de experimentar el evento en el tiempo t cuando todas las covariables son cero. Una estimación de este intercepto se obtiene usando el método de máxima verosimilitud lo cual no es posible usando la verosimilitud parcial de Cox.

Usando unos datos recolectados por la sección de registro y matrícula de la Universidad de Colombia, Sede Medellín durante los años 2005-2007 se ajustó un modelo estadístico de acuerdo a la ecuación (1) con el fin de identificar factores asociados con el número de semestres que transcurren hasta que un estudiante pierde la calidad de estudiante por primera vez. Por medio de un método *backward* de selección de variables, basado en el máximo valor- p , se identificó, del conjunto de covariables disponibles, aquellas que se asociaban de una manera estadísticamente significativa a este tiempo.

4. RESULTADOS

La tabla 1 muestra la significancia de los efectos de las covariables que el modelo identificó como significativamente asociadas al log-odds del número de semestres que transcurren hasta que un estudiante pierde la calidad de estudiante por primera vez. La variable semestre esta altamente asociada con el riesgo de perder la calidad de estudiante por primera vez (valor- $p < 0.0001$). De acuerdo a esta tabla, otras covariables asociadas con el log-odds son edad, puesto en el examen de admisión (donde el puesto es dividido por 10), género, estrato socioeconómico (alto o bajo), tipo de colegio (público, privado,

nocturno, otros), tipo de admisión (regular, mejor bachiller, municipio pobre e indígenas), tipo de tenencia de vivienda (propia, en arriendo o sin propiedad raíz), tiempo transcurrido desde grado de bachiller hasta admisión a universidad (menos de 2 años, 2 años, 3 años y más de 3 años).

Tabla 1. Significancia de factores que afectan la pérdida de calidad de estudiante

VARIABLE	GRADOS DE LIBERTAD	CHI-CUADRADO DE WALD	VALOR-P
Semestre académico	9	1637.8141	<.0001
Edad	1	18.5035	<.0001
Puesto en examen de admisión (Puesto*10 ⁻¹)	1	129.3390	<.0001
Estrato alto	1	15.3108	<.0001
Género femenino	1	102.0180	<.0001
Colegio Nocturno	1	15.2168	<.0001
Colegio Privado	1	19.9037	<.0001
Mejores Bachilleres	1	24.4675	<.0001
Municipios pobres e Indígenas	1	7.7204	0.0055
Sin Propiedad Raíz	1	9.1880	0.0024
Tiempo desde Bachiller = 2	1	14.0107	0.0002
Tiempo desde Bachiller = 3	1	7.4474	0.0064
Tiempo desde Bachiller > 3	1	18.6212	<.0001

La tabla 2 muestra las razones de odds junto con su respectivo intervalo de confianza del 95%.

Tabla 2. Razones de odds para factores que afectan la pérdida de calidad de estudiante

VARIABLE	RAZÓN DE ODDS	IC DEL 95%
Semestre 2 vs. Semestre 1	2.016	(1.874, 2.169)
Semestre 3 vs. Semestre 1	1.049	(0.956,1.151)

Semestre 4 vs. Semestre 1	0.439	(0.383, 0.503)
Semestre 5 vs. Semestre 1	0.359	(0.305, 0.422)
Semestre 6 vs. Semestre 1	0.245	(0.198, 0.303)
Semestre 7 vs. Semestre 1	0.148	(0.110, 0.200)
Semestre 8 vs. Semestre 1	0.098	(0.065, 0.149)
Semestre 9 vs. Semestre 1	0.084	(0.047, 0.148)
Semestre 10 vs. Semestre 1	0.030	(0.010, 0.092)
Edad	1.032	(1.017, 1.046)
Puesto en examen de admisión (Puesto*10 ⁻¹)	1.039	(1.032, 1.046)
Estrato alto	0.817	(0.739, 0.904)
Género femenino	0.712	(0.666, 0.760)
Colegio Nocturno	1.487	(1.218, 1.816)
Colegio Privado	0.861	(0.806, 0.920)
Mejores Bachilleres	0.233	(0.131, 0.415)
Municipios pobres e Indígenas	1.355	(1.094, 1.679)
Sin Propiedad Raíz	1.117	(1.040, 1.200)
Tiempo desde Bachiller = 2	1.204	(1.093, 1.327)
Tiempo desde Bachiller = 3	1.196	(1.052, 1.361)
Tiempo desde Bachiller > 3	1.298	(1.153, 1.461)

Observe que por ejemplo, estar en el semestre 9 disminuye el odds de perder la calidad de estudiante por primera en un 91.6% $((0.084-1)100\%=-91.6)$. Observe la tendencia decreciente del odds de perder la calidad de estudiantes por primera vez a lo largo de los semestres. También se observa que un incremento de un año en la edad incrementa el odds de perder la calidad de estudiante en un 3.2% mientras que subir diez puestos en el examen de admisión incrementa el odds en un 3.9%. Pertener al estrato alto disminuye este odds en un 18.3%. Ser mujer disminuye el odds en un 28.8%. Provenir de un colegio

nocturno incrementa el odds en un 48.7%, mientras que de un colegio privado disminuye el odds en un 13.9%. En cuanto al tipo de admisión se tiene que los admitidos como mejores bachilleres tienen un odds 76.7% menor, mientras que los admitidos mediante programas especiales para municipios pobres y comunidades indígenas tienen un incremento de los odds del 35.5%. El no tener propiedad raíz aumenta en un 11.7% el odds de perder la calidad de estudiante. Dejar pasar dos o tres años para ser admitido a la universidad, después de graduarse de bachiller incrementa el odds en un 20% aproximadamente, y dejar pasar un tiempo mayor aumenta el odds hasta un 29.8%.

5. CONCLUSIONES

Como puede observarse en los resultados se tiene un alto grado de concordancia entre las razones de odds y lo que se esperaría ocurriese según las variables que resultaron significativas.

Se identificaron y cuantificaron fenómenos que se sabe ocurren en el transcurrir la vida estudiantil universitaria, por ejemplo:

- El semestre académico tiene un efecto decreciente, es decir, a mayor el semestre que se está cursando menor son los odds de perder la calidad de estudiante.
- La calidad de los colegios privados disminuye el odds del evento.
- Los programas especiales de admisión para estudiantes de municipios pobres y comunidades indígenas (aunque garantizan la participación de estos sectores en la universidad) incrementan los odds de perder la calidad de estudiante.
- El tiempo desde el grado de bachiller hasta la admisión tiene un efecto creciente en los odds, es decir, los que dejan pasar 2 o más años tienen un odds mayor frente a los recién graduados o que esperan 1 año o menos para ser admitidos.

Como dirección futura se planea implementar un análisis similar a este pero usando datos recolectados desde la puesta en marcha de la Reforma Académica estipulada en el Acuerdo 033 de 2007. Esto

permitirá evaluar si las condiciones identificadas como causantes de la pérdida de la calidad de estudiante por primera vez en este análisis aún permanecen o si por el contrario existen nuevos factores que influyen en este evento.

AGRADECIMIENTOS

Los autores agradecen a María Zuly Zuluaga, del Centro de Computo de la Universidad Nacional de Colombia, Sede Medellín quien amablemente proporcionó los datos para realizar este análisis. También agradecen al profesor Juan Carlos Correa Morales de la Escuela de Estadística de la Universidad Nacional de Colombia, Sede Medellín por sus valiosos comentarios que enriquecieron este manuscrito. Trabajo patrocinado con recursos del proyecto DIME 20101007954.

REFERENCIAS

Allison P.D. *Survival analysis using the SAS system: a practical guide*. Cary (NC): SAS Institute Inc.; 1998

Castaño, E.A.; Gallón, S.; Gómez, K. y Vásquez, J. Análisis de los factores asociados a la deserción y graduación estudiantil universitaria, *Lecturas de Economía*, Vol. 65, pp. 9- 36, 2006.

Correa, J.C. y Lopera, C.M. Uso de tablas de vida para la estimación de la mortalidad estudiantil. Caso Universidad Nacional de Colombia, Sede Medellín, *Revista Universidad Eafit*, Vol. 44, Nro 150, pp. 46-50, 2008.

Roger Jesús Tovar y Juan Carlos Salazar. Un modelo lineal mixto adaptado a una cadena de Markov con espacio de estados ordinal. Aplicación a datos sobre promedios académicos de estudiantes. *Revista Colombiana de Estadística*. Diciembre 2009, volumen 32, no. 2, pp. 213 a 230

Woodward MW. *Epidemiology: study design and data analysis*. 2nd ed. Boca Raton (FL): Chapman and Hall/CRC Press; 2005