

Segmentación del Riesgo de Abandono de los Estudios Universitarios en la UPR-Cayey: Cohorte de Nuevo Ingreso de 2010

Por: Fernando Vázquez Calle* y José A. Cabrera Pérez**

enero de 2018

INTRODUCCIÓN

El año pasado el Departamento de Educación de EE. UU. (por sus siglas en inglés USDE) le aprobó a la UPR en Cayey una propuesta bajo el programa federal de Título V. Dicha propuesta está dirigida a estructurar un programa abarcador de retención, cuyo fin primordial es aumentar la tasa de graduación concentrando los esfuerzos y recursos en la prevención de la deserción. En los últimos diez (10) años la tasa de graduación general de la UPR en Cayey ha oscilado entre 40 y 49 por ciento.¹ Como parte del programa de retención diseñado, en la primera fase se requiere un modelo estadístico que permita identificar con anticipación los grupos de estudiantes de “alto riesgo” en abandonar los estudios sin completar el grado académico en el 150 por ciento del tiempo definido para los programas académicos vigentes. Estudios institucionales de la Oficina de Avalúo e Investigación Institucional (OPAII)² de la UPR-Cayey, han arrojado evidencia consistente y robusta de los factores de riesgo en completar el grado académico en la UPR-Cayey. Entre otros factores se pueden mencionar los siguientes: 1) índice académico de escuela superior, 2) puntajes en las pruebas de aptitud verbal y matemáticas del College Board, 3) sexo del estudiante, 4) tipo de escuela superior de procedencia (pública o privada), 5) ingresos del hogar, 6) generación del estudiante, 7) alternativa de admisión.

En respuesta a la necesidad del programa de retención que forma parte del proyecto de Título V, de contar con un instrumento que en forma objetiva y sistemática identifique temprana y preventivamente los estudiantes de mayor riesgo en fracasar, los investigadores de la OPAII desarrollaron un modelo estadístico de segmentación del riesgo en abandonar los estudios universitarios sin completar el grado académico. Dicho modelo está basado en una ecuación de regresión logística, mediante la cual se estimó el peso relativo de cada uno de los factores de riesgo en el evento binario de éxito o fracaso en completar el grado académico.

OBJETIVOS

Esta investigación tiene tres objetivos. Primero, evaluar tres modelos estadísticos de probabilidad (MEP) con el fin de seleccionar el que “mejor” vaticina la conducta de abandono de los estudios en la UPR-Cayey sin completar el grado de bachiller en un marco de tiempo de seis (6) años (12 semestres) consecutivos. Los tres modelos evaluados fueron: 1) regresión de árboles de decisión, 2) regresión logística y, 3) redes neuronales.

¹ Según datos oficiales suministrados por la Oficina de Avalúo e Investigación Institucional de la UPR en Cayey.

² Véase en el enlace <http://cayey.upr.edu/rectoria/oficina-de-planificacion-avaluo-e-investigacion-institucional-opaai/> dos estudios relacionados con este tema.

*Investigador Auxiliar en la Oficina de Avalúo e Investigación Institucional de la UPR en Cayey.

**Investigador en la Oficina de Avalúo e Investigación Institucional de la UPR en Cayey.

Segundo, basado en el MEP que mejor explica el riesgo de abandono de los estudios universitarios, pronosticar el riesgo de abandono de cada uno de los estudiantes de nuevo ingreso del año 2017. El programa de Título V utilizará estos resultados para identificar con anticipación los estudiantes de “alto riesgo” a los cuales se dirigirán los programas de intervención diseñados para mitigar la conducta de deserción.

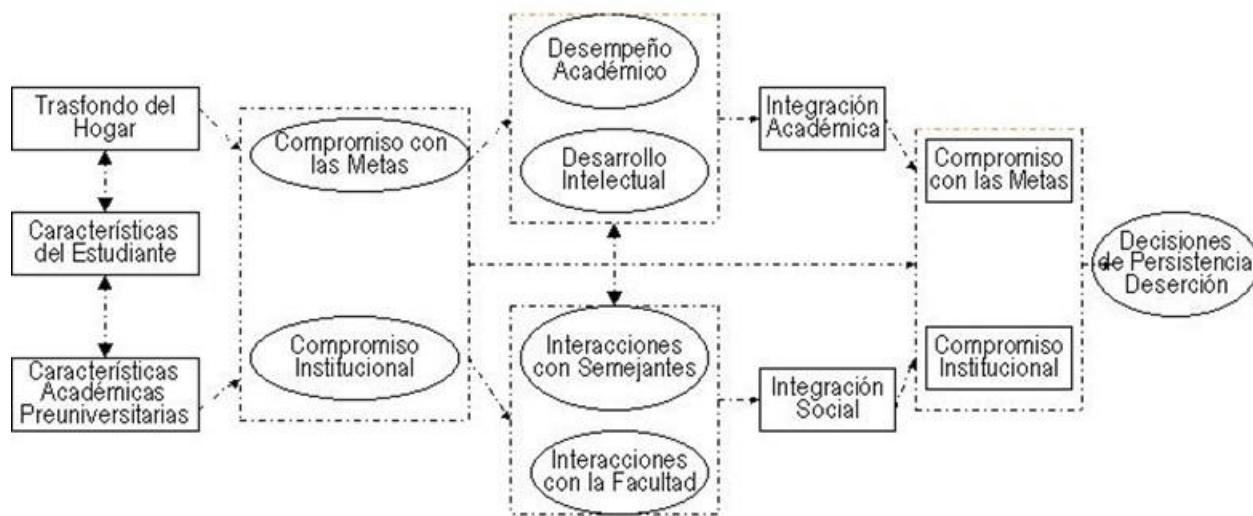
Tercero, segmentar los estudiantes de nuevo ingreso de 2010-2011 a base del riesgo de abandono y conocer el perfil sociodemográfico y académico preuniversitario de ellos.

REVISIÓN DE LITERATURA

A partir de los trabajos de Astin (1970), V. Tinto (1975, 1987 y 1993) y W. Spady (1971), uno de los temas más estudiados en el campo de la investigación institucional ha sido la conducta de persistencia/deserción. Hay abundante investigación acumulada en las universidades de EE. UU. y Europa sobre los factores que explican la conducta de persistencia/deserción. Puesto que el estudio que se presenta tiene como marco de referencia conceptual las teorías de los autores citados anteriormente, es conveniente comenzar con una breve descripción de dicho marco. El diagrama 1 muestra la trayectoria de correlaciones entre la decisión de persistencia/deserción y los distintos vectores de factores “causales” académicos y no académicos. Curiosamente este modelo conceptual, conocido como el modelo “interaccionista”, se inspiró en la teoría de suicidio del sociólogo E. Durkheim (1951), la cual suintamente postulaba que las deficiencias en la adaptación social del individuo aumentaban considerablemente el riesgo de suicidio. Brevemente, el modelo “interaccionista” afirma que el estudiante trae una serie de características individuales (trasfondo del hogar, psicológicas y de personalidad, académicas) que desarrollan un compromiso inicial con las metas. Posteriormente, en el ambiente universitario

dicho compromiso se modifica en función del grado de integración académica y social del

Diagrama 1. Modelo Interaccionista de Tinto



Elaboración propia basada en la teoría de V. Tinto.

estudiante al sistema social que representa la universidad. Finalmente, manteniendo constante las variables demográficas, académicas preuniversitarias y socioeconómicas, la decisión de persistencia/deserción es el resultado de cuán adecuada o inadecuada fue la integración académica y social.

Variables Académicas

En múltiples estudios sobre el fenómeno de la persistencia/deserción universitaria se ha encontrado una relación directa entre las variables académicas preuniversitarias (índice académico de escuela superior, percentil en escuela superior y puntaje en las pruebas estandarizadas de admisión universitaria) y el éxito en completar el grado académico. Entre otros, los siguientes autores han hallado evidencia empírica confirmando la relación significativa entre el éxito académico y las variables académicas preuniversitarias: Noble y Sawyer, 2002; Ishitani y Snider, 2004; Xiao, 1999; Gravely, 2003; Pascarella y Terenzini, 1980. Sin embargo, en un estudio de una universidad pública de cuatro años en EE. UU. (Zhu, 2002), no se halló evidencia de que el índice académico acumulado en escuela superior ni el puntaje total en el SAT (Scholastic Aptitude Test) tuvieran una influencia significativa en el resultado de graduarse o no.

Por otra parte, en el estudio anteriormente citado se concluyó que las variables con mayor capacidad predictiva eran el índice académico en el 1^{er} año de universidad y el número de cursos en que el estudiante fracasó. Gravely, 2003; Pascarella y Terenzini, 1980; Cabrera, La Nasa y Burkum, 2001, también encontraron que el índice académico del 1er. año de estudios universitarios tenía un efecto significativo sobre la probabilidad de graduarse. Encontraron que a mayor índice académico menor el riesgo de abandonar los estudios universitarios. Cabe mencionar que en los estudios de persistencia/deserción el índice académico de 1^{er} año de universidad es interpretado como un indicador de la integración académica del estudiante.

Variables No Académicas

En el estudio de Zhu (2002), de las variables demográficas consideradas (etnia, edad y género), sólo el género reflejó un efecto significativo en la probabilidad de éxito. Específicamente se encontró que las mujeres tenían una probabilidad de graduarse 31% mayor que los hombres. Otros autores que han encontrado resultados similares en cuanto al género son: Gravely, 2003; Cabrera, La Nasa y Burkum, 2001. La evidencia indicando que el género es un predictor del éxito en graduarse, aunque abundante ciertamente no ha sido unánime. Los siguientes autores han encontrado evidencia que contradice la anterior: Xia, 2000; Ishitani y Snider, 2004; Ishitani y DesJardins, 2002.

Una de las variables cuya relación con el éxito académico ha sido más estudiada es la que se conoce como la generación universitaria del estudiante. Aquellos estudiantes que provienen de un hogar en el que ninguno de sus padres asistió a la universidad se definen operacionalmente de 1^{ra} generación y, de 2^{da} generación aquellos que al menos uno de sus padres asistió a la universidad. También los resultados empíricos en cuanto al efecto de esta variable han sido mixtos. Whitley (1999), Elkins (1996), Pratt y Skaggs (1999), no hallaron que el status de la generación del estudiante influyera en la persistencia/deserción. Mientras que Duggan (2003), Suarez (1997), Arredondo (1999) y, Núñez y Cuccaro-Amin (1998), hallaron que el status de 1ra. generación reducía significativamente la probabilidad de éxito en completar el grado académico. Algunos investigadores han encontrado que los estudiantes de 1^{ra} generación tardan más tiempo en integrarse académica y socialmente a la comunidad universitaria (Terenzini, Rendon, et. al., 1994; Grayson, 1997). Esto a su vez, desde el punto de vista del enfoque “interaccionista” de Tinto, trae como resultado una menor probabilidad de persistir en el logro de las metas académicas. También, estudios de Terenzini, Springer, Yeager, Pascarella y Nora (1996) y, de York-Anderson y Bowman (1991) indican que los estudiantes de 1^{ra} generación perciben menos apoyo de sus padres que su contraparte de 2^{da} generación. También encontraron que los de 1^{ra} generación tenían inseguridad de sus destrezas académicas o que necesitaban cursos remediales. Cabe mencionar los interesantes hallazgos de Pratt y Skaggs, 1989; Terenzini, Springer, Yeager, Pascarella y Nora (1996) y; Burkhardt y Brown, 1999. Ellos encontraron que los estudiantes de 1^{ra} generación tenían aspiraciones académicas más bajas, así como expectativas de calificaciones más bajas, que los estudiantes de 2^{da} generación. En esa misma dirección Ishitani, Keving y Snider (2004), hallaron que las aspiraciones y expectativas de los padres con respecto a los planes de estudios universitarios de los hijos, influían negativamente sobre la probabilidad de éxito

académico de éstos. Relacionado con este aspecto del apoyo de los padres, Hossler, Schmit y Vesper (1999), encontraron que el involucramiento de los padres en los planes de estudio de los hijos estaba altamente correlacionado con un menor riesgo de deserción.

Duggan (2003), Plank y Jordan (2000), estudiaron el efecto de los siguientes factores del capital social del estudiante asociados con la escuela superior: tipo de escuela superior (pública o privada), localización (urbana o rural) y la rigurosidad del currículo. No hallaron que estos factores influyeran en la deserción universitaria.

Otra de las variables del trasfondo del estudiante señaladas en la literatura como un predictor de la persistencia/deserción es el ingreso del hogar. El ingreso es un indicador del acervo de riquezas. Ishitani y DesJardins (2002); Ishitani y Snider (2004), encontraron que el ingreso del hogar tenía un efecto significativo sobre la deserción. Mientras más bajos son los ingresos del hogar, más alto es el riesgo de abandono de los estudios. En un estudio longitudinal hecho con una base de datos de toda la nación de EE.UU. (NCES, 2000-194), Cabrera, La Nasa y Burkum (2001) encontraron una diferencia considerable por niveles socioeconómicos en las proporciones de los estudiantes de escuela superior que lograban obtener una grado universitario. Tan sólo un 13% de los estudiantes de nivel socioeconómico bajo habían obtenido un grado universitario. Mientras que el 57% de los estudiantes de nivel socioeconómico alto habían logrado un grado universitario.

La necesidad de dinero en efectivo para financiar los estudios universitarios esencialmente está determinada por los costos de la educación, por el acervo de riquezas de los padres y el estudiante (y la disposición de utilizar parte de dicho acervo para financiar la carrera universitaria) y, por las ayudas económicas disponibles en forma de becas, préstamos, exenciones y programas de estudio y trabajo. St. John, 1990; St. John, Andrieu, Oescher y Starkey, 1994; Stampen y Cabrera, 1986 y 1988, enfocan la persistencia y el éxito en graduarse como un producto derivado de las decisiones que se toman en el marco de la racionalidad económica. El estudiante persiste hasta graduarse en tanto y en cuanto los beneficios sociales y económicos (corriente de ingresos en la vida laboral promedio) de asistir a la universidad, son significativamente mayores que los costos y beneficios alternos (por ejemplo, trabajar a tiempo completo). De hecho, según los autores antes mencionados, la exención parcial o total de las tarifas de matrícula, las becas, los préstamos a bajo interés y los programas de estudio y trabajo, son incentivos para que el estudiante persista en la carrera universitaria, ya que buscan cerrar la brecha entre los costos de la educación y los beneficios de ir a la universidad.

Las investigaciones discutidas en el párrafo anterior comparten muchas de las ideas de lo que en el ámbito de las teorías explicativas de la persistencia en los estudios universitarios se ha denominado como el modelo economicista. Este modelo está predicado en la teoría de capital humano, cuyos precursores fueron Solow (1956), Shultz (1961) y Becker (1962 y 1964). Como resultado del desarrollo de la teoría de capital humano, la educación se comenzó a visualizar no sólo como un derecho, sino también como una inversión cuyos réditos el estudiante recibe durante toda su vida. La decisión de estudiar representa un sacrificio total o parcial de la corriente de ingresos durante el período que se estudia, así como también una erogación de

dinero en efectivo, con miras a capitalizar dicho valor invertido durante el resto de la vida del estudiante. Este enfoque se ha denominado como la teoría de capital humano, la cual en forma breve postula que la educación aumenta la productividad del ser humano, por lo cual es de esperarse que los ingresos sean mayores.

Para tener una idea concreta de los rendimientos de la inversión en capital humano, según el estudio de T. Mortenson (2000) en EE. UU., el ingreso de por vida de un hombre graduado de universidad, en promedio, resultó cerca de \$1.2 millones mayor que un hombre con diploma de escuela superior. En el caso de las mujeres la brecha fue sustancialmente menor (de \$600,000). En Puerto Rico, R. Cao-García (1988) comparó las tasas de retorno entre los graduados de universidad y los que sólo tenían diploma de escuela superior y encontró resultados similares. Halló que las tasas de retorno de las mujeres con grado universitario eran 33% mayor que las que sólo tenían diploma de escuela superior. Mientras que el retorno de la inversión de los hombres con grado universitario era 14% superior al de los hombres con diploma de escuela superior.³ Más reciente, los profesores J. I. Alameda-Lozada y A. González-Martínez (2018) estimaron que de 1999 a 2016 una persona con grado universitario (bachillerato, maestría o doctorado), en promedio, acumularía en su vida laboral y productiva, cerca de \$1.2 millones en salario, mientras que si hubiese tenido sólo escuela superior, hubiera acumulado alrededor de \$377 mil. En promedio, un graduado de escuela superior, durante su vida laboral acumula $\frac{1}{3}$ del salario que acumularía con un grado universitario.

MÉTODOS

Los modelos estadísticos empleados para estimar el riesgo de abandono de los estudios en la UPR-Cayey se basaron en los datos del cohorte de nuevo ingreso de agosto de 2010-11, el cual fue rastreado por seis años consecutivos (2010-11 a 2015-16) para identificar los estudiantes que no tuvieron éxito en completar el grado de bachiller. Dicho cohorte está compuesto 777 estudiantes procedentes de escuela superior que se matricularon en la universidad por primera vez. El archivo de matrícula de 2010 y el archivo compilado de grados otorgados de 2001 a 2016, preparados por la Oficina de Sistemas de Información de la UPR-Cayey, sirvieron de bases de datos para esta investigación. Como es lo usual, los archivos no tenían datos para algunas de las variables, por lo que los modelos estimados se basaron en una muestra de 654 estudiantes.

VARIABLES PREDICTORAS Y FUENTES DE INFORMACIÓN

Basado en investigaciones similares realizadas por la OPALL y la literatura pertinente, se incluyeron las siguientes variables como potenciales predictores del riesgo en abandonar los estudios sin completar el grado de bachiller.

1. Tipo de Escuela Superior (TES) – Variable cualitativa clasificada en tres categorías: 0=pública, 1=privada y 2=otra.

³ Esta asimetría tal vez es parte de la explicación del por qué la estructura de la matrícula en las instituciones de educación superior en Puerto Rico se ha ido “feminizando”.

2. Promedio Académico Acumulado en Escuela Superior (PAA-ES) – Es uno de los tres criterios (promedio académico de escuela superior, puntaje en las pruebas de aptitud verbal y matemáticas del College Board). Variable cuantitativa de 0.00 a 4.00.
3. Género (GEN) – Variable cualitativa, dónde 0=masculino, 1=femenino.
4. Generación Universitaria (GU) – Variable cualitativa. Se considera que un estudiante es de 2^{da} generación si al menos uno de los padres obtuvo un grado asociado o más. De otra manera se considera de 1^{ra} generación.
5. Ingreso del Hogar (IH) – En su origen es una variable cuantitativa. Sin embargo, como en la Solicitud Única de Admisión (SUA) de la UPR, que es la fuente de donde se obtuvieron los datos de esta variable, los datos se registran en grupos de ingresos mutuamente excluyentes y colectivamente exhaustivos, la variable se convierte en una de nivel ordinal. Recopilar datos de variables continuas en grupos ordinales mayormente se emplea cuando la información de interés de alguna forma resulta comprometedora para los respondientes. La conversión de variables continuas en variables nominales es una de las estrategias más utilizadas para lidiar con el problema de la no respuesta y la falta de precisión en los datos cuando se trata de información amenazante, comprometedora, embarazosa y socialmente reprobable. Aunque en la SUA esta variable tiene 12 grupos ordinales de ingreso, para propósitos de la estimación de los MEP evaluados, dichas categorías se fusionaron en tres: 2=alto (\$50,000 o más), 1=moderado (\$17,500 a \$49,999) y 0=bajo (menos de \$17,500).
6. Alternativas de Admisión (AA) – 0=2^{da} o 3^{ra} alternativa, 1=1^{ra} alternativa. El sistema de la UPR provee para que los aspirantes a ingresar puedan solicitar a un máximo de tres programas académicos (bien sea en un mismo recinto o recintos diferentes), como parte de los honorarios que pagan por solicitar admisión. La solicitud a los tres programas académicos se evalúa para la determinación de admisión o rechazo según el orden de prioridad definido en el formulario de solicitud.
7. Cursos Remediales en Español, Inglés y Matemáticas – Son variables binarias, tal que si el estudiante tomó el remedial se codificó igual a 1, y cero de cualquier otra manera.
8. Aptitud en Matemática (APT-MATE) – Puntaje en la parte de Aptitud Matemática de la prueba estandarizada de admisión universitaria administrada por el College Board. Es uno de los tres criterios que el sistema UPR emplea para aceptar o rechazar los estudiantes que solicitan admisión. Esta variable continua tiene un recorrido de 200 a 800.
9. Aptitud Verbal (APT-VERBAL) – Puntaje en la parte de Aptitud Verbal de la prueba estandarizada de admisión universitaria administrada por el College Board. Es otro de los tres criterios utilizados para decidir si un solicitante es aceptado o rechazado. Es una variable continua y tiene un recorrido de 200 a 800.

MODELOS ESTADÍSTICOS NO LINEALES DE PROBABILIDAD

La variable respuesta de este estudio es dicótoma. Por tanto, las técnicas de regresión lineal basadas en el algoritmo de los cuadrados mínimos ordinarios (CMO) están contra indicadas, porque los estimados de los efectos de las variables predictoras no son “óptimos” (Gujarati, 2009; Maddala, 2013; Kunt, Nachtsheim, Neter y Li, 2006). No obstante, en el campo de la estadística

se ha desarrollado un cúmulo de modelos no lineales de probabilidad, que permiten generar estimados “óptimos” en casos en que la variable dependiente sea binaria. Algunos de estos modelos son: 1) análisis discriminante, 2) probit, 3) regresión logística, 4) regresión de árboles de decisión y, 5) redes neuronales.⁴

El primero se descartó porque uno de sus supuestos requiere que todas las variables predictoras sean continuas; pero, algunos de los factores de riesgo que se tomaron en cuenta se miden en escalas no continuas (nominales y ordinales, tales como: sexo, alternativa de admisión, generación universitaria). Los modelos 2 y 3 generan estimados muy similares, ya que se trata de funciones del tipo sigmoide. El modelo 2 es más apropiado cuando se trata de diseños de experimentos, mientras que el modelo de regresión logística generalmente se aplica cuando los datos son observaciones no controladas. Considerando las ideas antes discutidas, para propósitos de este trabajo de investigación se seleccionaron los modelos 3, 4 y 5, los cuales se evaluaron en relación con su capacidad predictiva.

Regresión Logística

De los tres modelos evaluados el de regresión logística evidenció mayor poder predictivo, por lo que fue seleccionado como instrumento para estimar el riesgo de abandono de los estudios. Para medir la importancia relativa de las variables independientes sobre el riesgo de no completar el bachillerato, se estimó una ecuación de regresión logística. Dicha ecuación genera estimados óptimos del efecto neto de las variables independientes cuando la variable dependiente, como es el caso de este estudio, es dicótoma (Gujarati, 2009; Maddala, 2013; Kunt, Nachtsheim, Neter y Li, 2006; Aldrich y Nelson, 1986; Pampel, 2000; Demaris, 1992). La curva de esta ecuación es parecida a una S estilizada (véase el diagrama 2) y su campo de valores va de 0 a 1. Esta particularidad es lo que hace que los resultados se puedan interpretar como probabilidades de que el evento de interés ocurra. De ahí que la ecuación logística forma parte de la familia de modelos de probabilidad. En notación estadística la ecuación logística relacionando la variable dependiente con el grupo de variables independientes incluidas en el estudio se expresa así:

$$P(Y=0 \text{ ó } 1) = \text{EXP}(b_0+b_1X_1+b_2X_2+ \dots + b_kX_n+\varepsilon_i) \div [1+ \text{EXP}(b_0+b_1X_1+b_2X_2+ \dots + b_kX_n+\varepsilon_i)]$$

para toda $i = 1, 2, \dots, n$; $j = 1, 2, \dots, k$.

En el lado izquierdo de la ecuación se representa la variable dependiente (0= completó el bachillerato, 1=no completó el bachillerato) y en el lado derecho del signo de igualdad se

⁴ Para una explicación detallada de este grupo de técnicas estadísticas, pueden consultarse las siguientes referencias: 1) Kunt, M. H., Nachtsheim, C. J., Neter, J. y Li, W. (2006). *Applied linear statistical models*. 5th Edition. McGraw-Hill Company; 2) Gujarati y Porter (2009). *Basic econometrics*. 5th Edition. McGraw-Hill; 3) Baltagi, B. H. (2008). *Econometrics*. 4th Edition. Springer-Verlag; 3) Klecka, W. R. (1980). *Discriminant analysis*. Quantitative Applications in the Social Sciences Series. Thousand Oaks, Sage Publications; 4) Pampel, F. C. (2000). *Logistic regression*. Quantitative Applications in the Social Sciences Series. Thousand Oaks, Sage Publications; 5) Borooh, V. K. (2001). *Logit and probit*. Quantitative Applications in the Social Sciences Series. Thousand Oaks, Sage Publications; 6) Abdi, H., Valentin, D. y Edelman, B. (1998). *Neural networks*. Quantitative Applications in the Social Sciences Series. Thousand Oaks, Sage Publications.

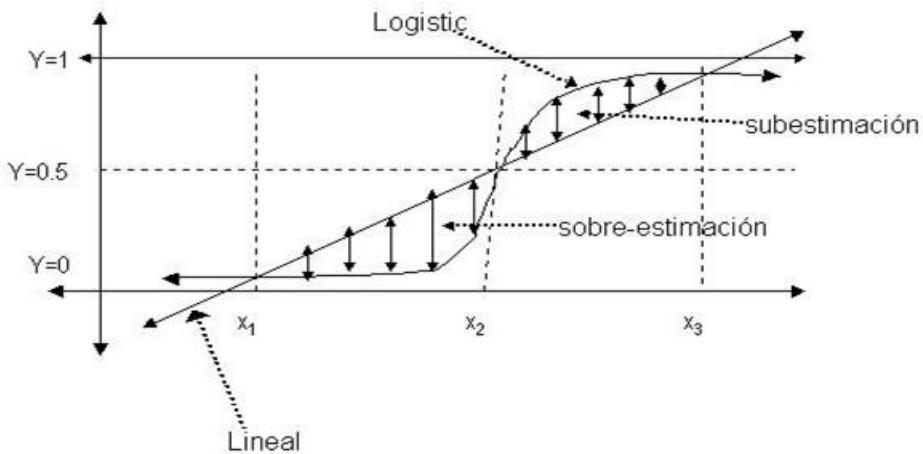
representan las variables predictoras (las X's), con sus respectivos coeficientes de regresión (b_0 , b_1 , b_2 , ..., b_k). El término ε_i se refiere a las desviaciones del modelo estimado, las cuales no se toman en cuenta porque en promedio tienden a cero. Para poder estimar la ecuación anterior por el algoritmo de máxima verosimilitud (MV),⁵ fue necesario hacer una transformación obteniendo los logaritmos naturales en ambos lados de la ecuación. El resultado de esta transformación es una ecuación lineal en los parámetros, que se expresa así:

$$\ln[P(Y=1)/ (1- P(Y=1))] = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_n$$

En el diagrama 2 se compara un modelo lineal de probabilidad con uno logístico. Como puede notarse a simple vista, en el arco entre x_1 y x_2 la ecuación lineal sobre estima la probabilidad de que ocurra el evento de interés, mientras que entre x_2 y x_3 ésta se subestima (Cabrera, 1994). También el modelo lineal puede arrojar resultados sin sentido, ya que el campo de valores de Y no necesariamente está restringido entre 0 y 1, como se espera de una variable binaria (Aldrich y Nelson, 1986). Como se muestra en el diagrama 2, cuando la variable dependiente es binaria el modelo lineal genera aproximaciones aceptables sólo en situaciones en que la distribución de los valores observados se concentra alrededor del centro (Goodman, 1977). Por las razones anteriores se seleccionó la transformación logística para modelar el éxito o fracaso del cohorte de nuevo ingreso de 2010 de la UPR-Cayey, dado el conjunto de variables predictoras especificado.

⁵ Es parecido al algoritmo de los CMO que se utiliza en las ecuaciones lineales de regresión, pero en lugar de minimizar las desviaciones cuadradas entre las observaciones y la línea de regresión, halla los valores de los coeficientes que maximizan la probabilidad de la realización de los valores observados de la variable dependiente. Para una explicación técnica consultese el texto de Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach*. 5th Edition. South-Western, Cengage Learning, Ohio, USA.

Diagrama 2. Modelos de Probabilidad Logistic y Lineal



Fuente: Elaboración propia basada en las explicaciones de los modelos lineales y no lineales en *Kunt, Nachtsheim, Neter y Li (2006)*.

Segmentación del Riesgo de Abandono

Mediante la adaptación del concepto de “prime prospect analysis (PPA)”, desarrollado en el campo de la investigación de mercados,⁶ se definieron cuatro (4) segmentos, los cuales se representan en el Diagrama 1. Como parte de la segmentación se elaboró un perfil comparativo de los cuatro (4) segmentos antes definidos. El método de segmentación PPA se basó en los estimados de riesgo de abandono generados por la modelo de regresión logística. Los segmentos se formaron mediante la clasificación conjunta del pronóstico de éxito o fracaso en completar el grado versus el éxito o fracaso observado.

⁶ Para una explicación detallada de los métodos de segmentación de mercados consultense las siguientes referencias: 1) Weinreich, N. K. (2011). *Hand-on social marketing: A step-by-step guide to designing change for good*. 2nd Edition, Sage Publications; 2) Myers, J. H. (1996). *Segmentation and positioning for strategic marketing decisions*. American Marketing Association, Chicago; USA. 3) Taq, J. (1997). *Multivariate analysis techniques in social sciences*. Thousand Oaks, Sage Publications.

Diagrama 1. Segmentación del Riesgo de Abandono de los Estudios Universitarios

Pronóstico de Riesgo de Abandono		
Riesgo Observado	Completó	Abandonó
Completó	Bajo Riesgo: Completaron y tenían perfil académico y sociodemográfico de bajo riesgo.	Vulnerables: Completaron y compartía características del perfil de alto riesgo
Abandonó	Riesgo Medio Bajo: Abandonaron y compartían características del perfil de bajo riesgo	Alto Riesgo: Abandonaron y tenían perfil de alto riesgo

Elaboración propia basada en el modelo de regresión logística estimado.

RESULTADOS

BONDAD DE AJUSTE DE LOS MEP ESTIMADOS

Los tres MEP especificados anteriormente fueron estimados. El modelo de regresión logística evidenció la tasa de pronósticos correctos más alta (67% versus 60% y 63%). Además, la ecuación de regresión logística reflejó un ajuste aceptable, ya que los pseudo coeficientes de determinación R^2 de Conteo y McKelvey-Zavoina⁷, arrojaron valores de 67% y 68%, respectivamente. Es oportuno indicar que se excluyeron algunos predictores que, en estudios realizados por la OPAII, han evidenciado una robusta contribución explicando la conducta de abandono (entre otros: GPA en el 1er año de universidad, repeticiones de cursos, abandono temporal de los estudios). Sin duda, si los predictores excluidos se hubieran tomado en cuenta el modelo de regresión logística hubiese registrado un poder explicativo mucho mayor. No obstante, como el objetivo principal de esta investigación es generar una ecuación que sirva de instrumento para identificar tempranamente (desde antes de matricularse por primera vez) a los estudiantes admitidos, que, por su perfil, pueden considerarse de “alto riesgo”, se excluyó cualquier variable cuyos datos no están disponibles previo a iniciar los estudios universitarios.

⁷ Ambos estadísticos de la bondad de ajuste son similares al coeficiente determinación del modelo lineal estadístico, el cual mide el porcentaje de la variación en la variable dependiente explicado por las variaciones en el grupo de variables predictoras.

SIGNOS ESPERADOS DE LOS PREDICTORES

Los signos algebraicos que toman los predictores son indicativos de la dirección de la relación entre la variación de los valores de los predictores y la variación de la variable dependiente. En la ecuación de regresión estimada se postuló que la variable dependiente (Y) era el resultado de abandonar o completar el grado (0=completó el grado, 1=abandonó sin completar el grado). Se definió de esta manera porque se interesaba estimar el riesgo de abandonar los estudios sin completar el grado académico. Para entender la dirección de la relación entre los predictores (X's) y la Y (y por ende, el signo de cada coeficiente), es indispensable conocer la codificación empleada tanto para las X's como para la Y. Se espera que el signo de los coeficientes de las variables cuantitativas Aptitud Verbal, Aptitud en Matemáticas y GPA de Escuela Superior sea negativo. Un signo negativo indicando una relación inversa entre el predictor y la Y, quiere decir que cuando el valor de X disminuye el riesgo de abandonar los estudios aumenta.

Tabla 1. Signo Esperado de los Predictores y Categorías de Control de las Variables Cualitativas

Predictores	Signo Esperado	Categoría de Control	Relación con el Riesgo de Abandono
Generación Universitaria	Positivo	2 ^{da} Generación	1 ^{ra} generación tiene mayor riesgo de abandono
Ingreso del Hogar	Positivo	Alto (\$50,000 o más)	Menor ingreso mayor riesgo de abandono
Alternativa de Admisión	Positivo	1 ^{ra} Alternativa	2 ^{da} alternativa tienen más riesgo de abandono
Género	Positivo	Femenino	Hombres tienen más riesgo de abandono
Tipo de Escuela Superior	Positivo	Privada	Escuela pública tiene mayor riesgo de abandono
Aptitud en Verbal	Negativo	No Aplica	A menor puntaje, mayor riesgo de abandono
Aptitud en Matemáticas	Negativo	No Aplica	A menor puntaje, mayor riesgo de abandono
GPA-Escuela Superior	Negativo	No Aplica	A menor puntaje, mayor riesgo de abandono
Remedial en Matemática		Sí Tomó el Remedial	
Remedial en Español		Sí Tomó el Remedial	
Remedial en Inglés		Sí Tomó el Remedial	

En el caso de las variables cualitativas incluidas en el modelo de regresión es indispensable conocer la categoría que se definió como referencia o control. La Tabla 1 resume los signos esperados de los predictores y las categorías de control definidas. Exceptuando las variables binarias relativas a los cursos remediales que se ofrecen en la UPR-Cayey (español, matemáticas e inglés), se espera que las otras variables cualitativas registren una relación directa. Como puede verse en la tabla antes mencionada, no se especificó el signo de los remediales en español, inglés y matemáticas. Por varias razones no resulta del todo claro la dirección de la relación de los remediales y el riesgo de abandono. De acuerdo con nuestra interpretación, los remediales podrían influir por dos canales. Uno, mejorando las destrezas en las materias y; dos, facilitando la adaptación del estudiante al sistema social y académico de la universidad, la cual a veces resulta traumática. A primera vista, por sentido común se puede plantear que el tomar los remediales reduce el riesgo de abandono. El problema que tiene tal premisa es que los que no toman los remediales (y como resultado no se pueden considerar en el modelo de segmentación del riesgo de abandono), precisamente son los que por haber obtenido puntajes en las pruebas de admisión universitaria por encima de los puntos de cortes definidos por la UPR-Cayey, no

necesitan tomar los remediales porque, según el desempeño en dichas pruebas, están preparados para enfrentar el rigor de los cursos universitarios.

Poder Explicativo de los Predictores

En los modelos de regresión lineal la contribución de cada predictor (X) explicando los cambios en la variable dependiente (Y), se mide a través de los coeficientes estimados (b), haciendo la famosa suposición de *ceteris paribus* (Gujarati y Porter, 2009). Esto es, manteniendo todo lo demás igual (*ceteris paribus*), el coeficiente b mide, en promedio, el cambio en Y explicado por cada unidad de cambio en X. Sin embargo, debido a que el modelo de regresión logística no es lineal, sino que mediante transformaciones algebraicas se linealiza en los parámetros, los coeficientes de regresión (b) no transmiten en forma clara los cambios en Y asociados a los cambios en cada predictor X. Por tal razón, en los modelos de regresión logística la interpretación del efecto de cada predictor sobre Y, está basada en el estadístico conocido como "odds ratio" (OR).⁸ Algunos programados estadísticos en lugar de presentar el término "odds ratio" o sus siglas OR, muestran la expresión e^b , donde la base es la constante matemática e ($e \approx 2.7183$) y b el coeficiente de regresión estimado de un predictor X dado. La expresión e^b se usa porque el OR de un predictor en una ecuación de regresión logística (la cual se ha linealizado calculando el logaritmo natural en ambos lados de la ecuación) se obtiene utilizando la e como base y el coeficiente de regresión (b) como exponente.

Efectos de las Variables Predictoras

La Tabla 2 muestra los estimados de la ecuación de regresión formulada anteriormente. De los tres predictores de escala cuantitativa incluidos (APT-VERB, APT-MATE y PAA-ES), sólo el

⁸ El concepto de OR se acuñó de las apuestas en los juegos de azar y en español se ha traducido mayormente como oportunidades relativas. En notación los "odds ratio" se expresan así:

$$OR = (E_{\text{éxito}_{\text{expuestos}}} / S - E_{\text{éxito}_{\text{expuestos}}}) \div (E_{\text{éxito}_{\text{no-expuestos}}} / S - E_{\text{éxito}_{\text{no-expuestos}}})$$

donde:

$E_{\text{éxito}_{\text{expuestos}}}$ = número de casos del grupo expuesto al factor de riesgo en que ocurrió el evento de interés (éxito)

$E_{\text{éxito}_{\text{no-expuestos}}}$ = número de casos del grupo no expuesto en que ocurrió el evento de interés (no-éxito)

S = espacio muestral o total de observaciones

Para entender este concepto es mejor ilustrarlo con un ejemplo. Suponiendo lo siguiente en el cohorte de nuevo ingreso del año t) 562 eran escuela pública y 307 de las privadas, 2) de los que procedían de escuela pública 312 no completaron el grado y de las privadas 146 no lo completaron. Se podría preguntar, ¿por cuánto aumenta el riesgo de fracaso dado el tipo de escuela superior de procedencia? Para estimar dicho factor de riesgo se puede emplear el estadístico OR. Según la fórmula de OR, el aumento en el factor de riesgo está dado por:

OR = $[312/(869-312)]/[146/(869-146)] = (312/557)/(146/723) = 0.56/0.21 = 2.67$. Por tanto, el riesgo de fracaso de los estudiantes de escuela pública es un poco más de 2½ veces mayor que los de escuela privada. Nótese que el OR no es otra cosa que el cociente entre la oportunidad que ocurre el evento de interés en el grupo expuesto al factor de riesgo y la oportunidad en el grupo no expuesto al factor de riesgo.

promedio académico de escuela superior (PAA-ES) arrojó evidencia contundente de influir significativamente en el riesgo abandonar los estudios. Tal como se esperaba, el signo del promedio académico de escuela superior es negativo, indicando que, en promedio y aislando el efecto de los otros predictores, a menor promedio académico, mayor el riesgo de abandonar los estudios universitarios. El OR del PAA-ES es igual a 0.257, lo cual quiere decir, que manteniendo todo lo demás constante, en promedio, por cada décimo (0.10) que baja el PAA-ES, el riesgo de abandonar los estudios aumenta en 7.4%. Obsérvese en la tabla antes mencionada que las variables APT-VERBAL y APT-MATE no ejercen influencia estadísticamente significativa en el riesgo de abandonar los estudios.

Los resultados empíricos de la regresión estimada confirmaron los signos esperados de los coeficientes (b) de todas las variables predictoras. Como puede verse en la Tabla 2, los grupos de ingreso medio y bajo arrojaron un valor-p menor que 0.05, indicativo de que el ingreso ejerce una influencia significativa sobre el riesgo de abandonar los estudios. La evidencia que arrojó la ecuación de regresión logística confirmó la relación inversa entre ingreso y riesgo de abandonar los estudios. Como categoría o grupo de control se definió a los estudiantes provenientes de hogares con ingresos más altos (\$50 mil o más). Por tanto, el estadístico e^b se interpreta tomando como base el grupo de control definido. El OR de los grupos de ingreso medio y bajo es igual a 1.96 y 1.95, respectivamente. Esto quiere decir, que los estudiantes procedentes de hogares con ingreso medio y bajo tienen un riesgo de casi dos veces mayor que los estudiantes del grupo control de ingresos altos. La procedencia de escuela superior influye significativamente en el riesgo de abandono de los estudios. Los de escuela pública tienen un riesgo de aproximadamente 150% (OR=1.589) mayor que los de escuela privada. Asimismo, los estudiantes del género masculino tienen un riesgo de abandonar los estudios significativamente mayor, alrededor de dos veces (OR=2.15) el riesgo de las féminas. También los estudiantes que no fueron admitidos en 1^{ra} alternativa tienen un riesgo de abandonar cerca de dos veces el riesgo de los que fueron admitidos en 1^{ra} alternativa.

Tabla 2. Coeficientes de la Regresión Logística Estimados

Variables Predictoras	Valor				
	b	Error Estándar	de Wald	Valor P	e ^b
Generación Universitaria (2 ^{da}) ^a	-0.026	0.192	0.018	0.894	0.975
Ingreso del Hogar (Alto: \$50,000 ⁺) ^a			8.178	0.017	
Medio (\$17,500-\$49,999)	0.528	0.271	3.808	0.051	*** 1.696
Bajo (Menos de \$17,500)	0.669	0.234	8.177	0.004	* 1.951
Alternativa de Admisión (1 ^{ra}) ^a	0.655	0.228	8.269	0.004	* 1.926
Género (femenino) ^a	0.765	0.193	15.675	0.000	* 2.148
Tipo de Escuela Superior (privada) ^a	0.456	0.200	5.228	0.022	** 1.578
Aptitud en Verbal	-0.002	0.001	2.384	0.123	0.998
Aptitud en Matemáticas	0.001	0.001	0.891	0.345	1.001
GPA-Escuela Superior	-1.359	0.305	19.857	0.000	* 0.257
Remedial en Español (Sí) ^a	0.258	0.231	1.250	0.263	1.294
Remedial en Inglés (Sí) ^a	0.464	0.237	3.840	0.050	** 1.590
Remedial en Matemáticas (Sí) ^a	-0.864	0.255	11.477	0.001	* 0.421
Constante	4.031	1.464	7.583	0.006	* 56.336

* Significación estadística igual o menor que 0.01.

** Significación estadística entre 0.01 y 0.05.

*** Significación estadística entre 0.05 y 0.10.

^a Dentro del paréntesis se define la categoría de control o de referencia.

Fuente: Elaboración propia (2018).

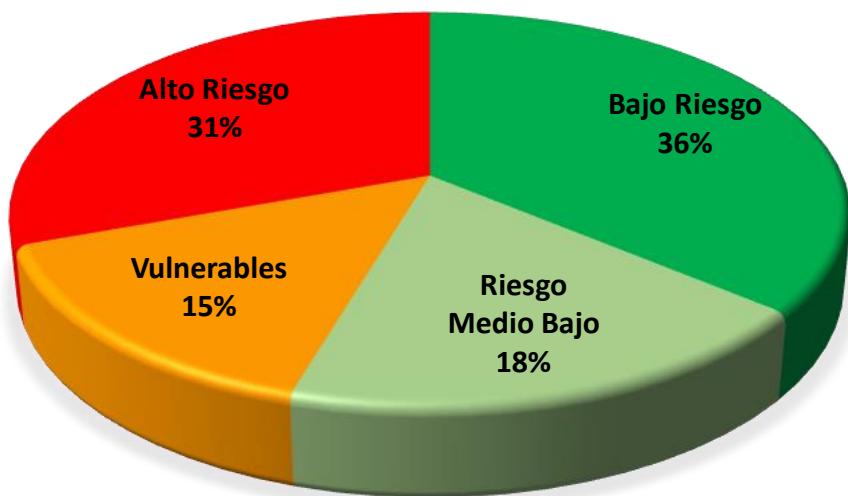
Por último, dos de los tres cursos remediales, en particular matemáticas e inglés, arrojaron evidencia suficiente para concluir que ejercen una influencia estadísticamente significativa (valor-p = 0.001 y 0.050, respectivamente). Los signos de los remediales de matemáticas e inglés fueron negativo y positivo, respectivamente. Se reitera que la relación entre el riesgo de abandono de los estudios y tomar los remediales es confusa. Por tanto, la interpretación de los estimados debe tomarse con suma cautela.

PERFIL DE LOS SEGMENTOS

El tamaño relativo de los segmentos, definidos en función del riesgo de abandono de los estudios, se muestra en la Gráfica 1. Se destaca que los dos segmentos que representan los extremos opuestos, Bajo Riesgo y Alto Riesgo, cada uno comprende alrededor de una tercera parte de todos los estudiantes de nuevo ingreso. En cambio, los segmentos del centro, Vulnerables y Riesgo Medio Bajo, representan 15% y 18%, respectivamente.

La Tabla 3 resume el contraste de las características académicas preuniversitarias y sociodemográficas entre los cuatro segmentos en los cuales se clasificaron los estudiantes de

Gráfica 1
Tamaño de los Segmentos



nuevo ingreso. Los descriptores de escala cuantitativa muestran el valor promedio de cada segmento. Mientras que las variables cualitativas muestran los porcentajes, los cuales se calcularon tomando como base el total de cada categoría. La primera variable descriptora que se muestra en la mencionada tabla es el pronóstico de riesgo de abandono de los estudios, calculado a partir de la ecuación de regresión logística estimada. Obsérvese que el riesgo de abandono entre los primeros dos segmentos, Bajo Riesgo y Riesgo Medio Bajo, es comparativamente pequeño y muy similar en ambos segmentos. Respectivamente, el riesgo de abandono de los segmentos mencionados fue 35.2% 38.8%, lo cual quiere decir, que en promedio, alrededor de 1 de cada 3 estudiantes, clasificados en tales segmentos, no completa el grado académico en el 150% del tiempo definido. Si se visualiza el complemento del riesgo de abandono, se espera que aproximadamente 2/3 partes los estudiantes de los segmentos antes mencionados, complete el grado académico. Mientras que en los otros dos segmentos, Vulnerables y Alto Riesgo, la expectativa es que alrededor de una tercera parte (1/3) complete el grado.

Nótese que las variables descriptoras que miden características académicas preuniversitarias siguen el patrón de que los segmentos de menor riesgo (Bajo Riesgo y Riesgo Medio Bajo) reflejan valores típicos más altos, evidenciando un desempeño académico superior a los estudiantes de los segmentos que tienen mayor riesgo de abandono (Alto Riesgo y Vulnerables). Resulta curioso que el puntaje promedio en la prueba de aprovechamiento en inglés es muy similar en los cuatro segmentos. Aunque los dos segmentos de menor riesgo registraron un puntaje promedio mayor, la diferencia es bastante pequeña. Esto significa que el aprovechamiento en inglés, medido a través de la prueba de admisión universitaria del College Board, tiene un bajo poder para discriminar entre los segmentos.

Tabla 3. Perfil Académico y Sociodemográfico por Segmento de Riesgo

Descriptor	Bajo Riesgo	Riesgo Medio Bajo	Vulnerables	Alto Riesgo
Pronóstico de Riesgo	35.2%	38.8%	62.4%	65.3%
Aprovechamiento en Español	561	559	539	533
Aprovechamiento en Inglés	540	540	535	529
Aprovechamiento en Matemáticas	594	586	559	559
Aptitud en Matemáticas	584	578	563	569
Aptitud en Verbal	582	576	552	555
Promedio Académico				
Escuela Superior	3.85	3.81	3.58	3.55
Índice General Solicitud (IGS)	318	314	297	296
Escuela Superior				
Pública	32.3%	18.3%	16.6%	32.8%
Privada	42.4%	18.8%	11.8%	26.9%
Generación Universitaria				
1 ^{ra} Generación	32.0%	17.0%	16.6%	34.4%
2 ^{da} Generación	38.6%	19.4%	13.8%	28.3%
Género				
Femenino	45.5%	22.9%	11.1%	20.6%
Masculino	16.5%	9.4%	22.6%	51.4%
Alternativa de Admisión				
1 ^{ra} Alternativa	41.0%	21.0%	12.7%	25.3%
2 ^{da} ó 3 ^{ra} Alternativa	12.5%	6.3%	25.0%	56.3%

Del grupo de nuevo ingreso procedente de escuela pública, cerca de una tercera parte se clasificó en los segmentos de Bajo Riesgo y Alto Riesgo. Merece resaltarse que la procedencia de escuela pública de los segmentos antes mencionados casi duplica la de los otros dos segmentos (Riesgo Medio Bajo y Vulnerables). Un comportamiento parecido se observó en la variable Generación Universitaria. Como puede verse en la Tabla 3, también cerca de una tercera parte de los estudiantes de 1^{ra} generación están clasificados en los segmentos de Bajo Riesgo y Alto Riesgo. Los estudiantes del sexo femenino tienen una presencia sustancialmente mayor en el segmento de Bajo Riesgo (45.5%). En cambio, la presencia de varones en el segmento de Alto Riesgo es proporcionalmente muchísimo más alta que en los otros segmentos. Algo similar se observó con los estudiantes que fueron admitidos en 2^{da} o 3^{ra} alternativa; de todos los estudiantes que no fueron admitidos en 1^{ra} alternativa, el 56.3% clasificó en el segmento de Alto Riesgo.

CONCLUSIONES

De acuerdo con la ecuación de regresión logística estimada, en orden de importancia, se halló que las siguientes características de los estudiantes de nuevo ingreso son factores que contribuyen significativamente en el riesgo de abandonar los estudios universitarios.

1. Promedio Académico Acumulado en Escuela Superior
2. Género
3. Alternativa de Admisión
4. Tipo de Escuela Superior de Procedencia
5. Ingreso del Hogar
6. Remediales de Matemáticas e Inglés

Los puntajes en las pruebas de aptitud en matemáticas y verbal no arrojaron evidencia de ser factores de riesgo significativo. Sin embargo, en un estudio realizado por F. Vázquez-Calle y J. A. Cabrera-Pérez (2005)⁹ ambas pruebas evidenciaron una influencia significativa sobre el promedio académico en el 1^{er} año de universidad. Considerando los resultados del estudio antes citado, parece ser que las pruebas de aptitud del College Board ejercen una influencia indirecta sobre el riesgo de abandonar los estudios.

REFERENCIAS

Alameda-Lozada, J. I. y González-Martínez, A. (2015). *Del government right sizing al government social right sizing: La UPR y su rol imprescindible para el país*. Plan Sostenible para la UPR 2018.

Aldrich, J. H. y Nelson, F. D. (1984). *Linear probability, logit, and probit models*. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-045. Thousand Oaks, CA: Sage.

Arredondo, M. (1999). *First generation college students at a selective, four-year institution: Transition to college, adjustment in college and self-image*. Unpublished doctoral dissertation, University of California at Los Angeles.

Astin, A. W. (1975). *Preventing students from dropping out*. San Francisco: Jossey-Bass.

Baltagi, B. H. (2008). *Econometrics*. 4th Edition. Springer-Verlag.

Burkhardt, R. L. y Brown, H. E. (1999). *Predicting academic success of college students: Does first generation status matter?* Paper presented at the Association for Institutional Research Annual Forum, Seattle, Washington.

Cabrera, A. F., Castaneda, M. B., Nora, A. y Hengstler, D. (1992). *The convergence between two theories of college persistence*. Journal of Higher Education 63, 143-164.

⁹ Puede accederse en el enlace <http://cayey.upr.edu/rectoria/oficina-planificacion-avaluo-e-investigacion-institucional-opaii/>.

Cabrera, A. F., La Nasa, S. M. y Burkum, K. R. (2001). *Pathways to a four-year degree: The higher education story of one generation*. Presented at the Association for Institutional Research Annual Forum.

Cao-García, R. J. (1988). *Tasas de rendimiento de la inversión en educación*. Unidad de Investigaciones Económicas, Recinto de Río Piedras de la Universidad de Puerto Rico.

Chase, P. B. (2000). *First generation college freshmen: The impact of academic support on retention, grade point average, and locus of control*. In M. B. Duggan (2003). *The effect of social capital on the first-year persistence of first generation college students*. Paper presented at the Association for Institutional Annual Forum, Tampa, Florida.

Duggan, M. B. (2003). *The effect of social capital on the first-year persistence of first generation college students*. Paper presented at Association for Institutional Research Annual Forum, Tampa, Florida.

Elkins, S. K. (1996). *The ties that bind: A study of persistence of first-generation college students in a 4-year institution*, in M. B. Duggan (2003). Paper presented at Association for Institutional Research Annual Forum, Tampa, Florida.

Gómez-Alvis, C., Cano-Gamboa, C., Zuluaga-Díaz, F. y Cardona-Acevedo, M. *Diferencias y similitudes en las teorías del crecimiento económico*. Departamento de Economía, Escuela de Administración, Universidad EAFIT.

Gravely, A. R. (2003). *Using logistic regression to assess the impact of first-year experience programs on student retention*. Paper presented at the Association for Institutional Research, Tampa, Florida.

Grayson, J. P. (1997). *Academic achievement of first-generation students in a Canadian university*. Research in Higher Education, 38, 659-673.

Gujarati, D. M. y Porter, D. C. (2009). *Basic econometrics*. 5^{ta} edición. McGraw-Hills/Irwin, Inc.

Hossler, D., Schmit, J. y Vesper, N. (1999). *Going to college: How social, economic, and educational factors influence the decisions students make*. Hopkins University Press, Baltimore, MD.

Ishitani, T. y DesJardins, S. L. (2002). *A longitudinal investigation of dropout from colleges in the United States*. Journal of College Student Retention, 4(2), 173-201.

Ishitani, T. y Snider, K. G. (2004). *Longitudinal effects of college preparation programs on college retention*. Paper presented at the Association for Institutional Research Annual Forum, Boston, Massachusetts.

Nieves Piovani, J. E., Leguizamón, M. y Steclow, U. (enero-marzo 2015). *Explorando el impacto político de la WEB con técnicas avanzadas de análisis*. Revista Especializada en Periodismo y Comunicación, vol. 1 núm. 45.

Klecka, W. R. (1980). *Discriminant analysis*. Quantitative Applications in the Social Sciences Series. Thousand Oaks, Sage Publications.

Kunt, M. H., Nachtsheim, C. J., Neter, J. y Li, W. (2006). *Applied linear statistical models*. 5th Edition. McGraw-Hill Company.

Menard, S. (2010). *Logistic regression: From introductory to advanced*. Thousand Oaks, Sage Publications.

Mortenson, T. (2000). *Private economic benefit/cost ratios of a college investment for men and women 1967 to 1999*. Postsecondary Education Opportunity, 101, 1-15.

Myers, J. H. (1996). *Segmentation and positioning for strategic marketing decisions*. American Marketing Association, Chicago; USA.

Noble, J. y Sawyer, R. (2002). *Accuracy of high school grades and college admissions test scores for predicting different levels of academic achievement in college*. Paper presented at the Association for Institutional Research Annual Forum, Toronto, Canada.

Nunez, A. M. y Cuccaro-Amin, S. (1998). *First generation students: undergraduates whose parents never enrolled in postsecondary education*. U.S. Department of Education: Washington, D.C.

Pascarella, E. y Terenzini, P. (1983). *Predicting voluntary freshman year persistence/withdrawal behavior in a residential university: an analytic validation of the Tinto model*. Journal of Educational Psychology 85: 215-226.

Pratt, P. A. y Skaggs, C. T. (1989). *First generation college students: Are they at greater risk for attrition than their peers?* Research In Rural Education, 6, 31-34.

Solow, R. (1956). *A contribution to the theory of economic growth*. Quarterly Journal of Economics, febrero de 1956.

Spady, W. G. (1970). *Dropouts from higher education: An interdisciplinary review and synthesis*, Interchange, 1, 109-121.

Taq, J. (1997). *Multivariate analysis techniques in social sciences*. Thousand Oaks, Sage Publications.

Terenzini, P. T., Rendon, L. et al. (1994). *The transition to college: diverse students, diverse stories*. Research in Higher Education, 35,58-73.

Tinto, V. (1975). *Dropout from higher education. A theoretical synthesis of recent research*. Review of Educational Research, 45, 89-125.

Tinto, V. (1987). *Leaving college: rethinking the causes and cures of student attrition*. University of Chicago Press: Chicago, IL.

Weinreich, N. K. (2011). *Hand-on social marketing: A step-by-step guide to designing change for good*. 2nd Edition, Sage Publicatio

Whitley, S. M. (1999). *A comparative study of the persistence of first-generation college students and other undergraduates*. In M. B. Duggan (2003). *The effect of social capital on the first-year persistence of first generation college students*, Paper presented at Association for Institutional Research, Tampa, Florida.

Xiao, B. (1999). *The impact of freshman year academic success on student persistence and bachelor's degree completion*. Paper presented at the Association for Institutional Research Annual Forum, Seattle, Washington.

York-Anderson, D. y Bowman, S. L. (1991). *Assessing the college knowledge of first-generation and second-generation college students*. Journal of College Student Development, 32, 116-121.

Zhu, L. (2002). *How the first-year college experience contributes to persistence*. Paper presented at the Association for Institutional Research Annual Forum, Toronto, Canada.