

UNIVERSIDAD DE COSTA RICA
FACULTAD DE CIENCIAS ECONÓMICAS
ESCUELA DE ECONOMÍA

Efectos socio-demográficos en el rendimiento académico de los estudiantes de carreras con
alta carga matemática: Universidad de Costa Rica 2009-2013

Memoria de Seminario de Graduación para optar por el Grado Académico de Licenciatura en
Economía

Marcos David Castillo Bastos

B01440

Ciudad Universitaria Rodrigo Facio

San José, Costa Rica

Agosto, 2016

UNIVERSIDAD DE COSTA RICA
FACULTAD DE CIENCIAS ECONÓMICAS

Acta #_09-08-16_

Acta de la Sesión _09-08-16_ del Comité Evaluador de la Escuela de _Economía_, celebrada _viernes 12 de agosto del 2016, con el fin de proceder a la Defensa del Trabajo Final de Graduación de _ **Marcos Castillo Bastos**_, carné _**B01440**_, quien optó por la modalidad de: _Seminario de Graduación_.
Presentes: _M.Sc. Max Alberto Soto Jiménez_, quien presidió; _Dra. Yanira Xinirachs Salazar_, como tutora; _Dr. Leonardo Garnier Rímolo_, como lector, _Dr. Juan Rafael Vargas_, como lector y _Dr. Pablo Sauma Fiat_, quien actuó como secretario de la sesión.

Artículo 1

El Presidente informa que el expediente del estudiante postulante contiene todos los documentos que el Reglamento exige. Declara que ha cumplido con todos los requisitos del Programa de la Carrera de Licenciatura en _**Economía**_.

Artículo 2

El estudiante hizo la exposición del Trabajo Final titulado "**Efectos socio-demográficos en el rendimiento académico de los estudiantes de carrera con alta carga matemática: Universidad de Costa Rica 2009-2013**".

Artículo 3

Terminada la disertación, los miembros del Comité Evaluador, interrogan al postulante el tiempo reglamentario. Las respuestas fueron Satisfactorias, en opinión del Comité.
(satisfactorias/insatisfactorias)

Artículo 4

Concluido el interrogatorio, el Tribunal procedió a deliberar.

Artículo 5


Efectuada la votación, el Comité Evaluador consideró el Trabajo Final de Graduación Satisfactorio, y lo declaró aprobado.
(Satisfactorio/insatisfactorio) (aprobado/no aprobado)

Artículo 6


El presidente del Comité Evaluador comunicó en público al aspirante, el resultado de la deliberación y lo declaró Licenciado en Economía.

Se le indicó la obligación de presentarse al Acto Público de Juramentación. Luego se dio lectura al acta que firmaron los miembros del Comité y el estudiante a las 14:30 horas.


M.Sc. Max Alberto Soto Jiménez
Director a.i. de la Escuela


Marcos Castillo Bastos
B01440


Dra. Yanira Xinirachs Salazar
Tutora del Trabajo


Dr. Leonardo Garnier Rímolo
Lector del Trabajo


Dr. Juan Rafael Vargas Brenes
Lector del Trabajo


Dr. Pablo Sauma Fiat
Secretario de la Sesión

Según lo establecido en el Reglamento de Trabajos Finales de Graduación, artículo 39 "... En caso de trabajos sobresalientes; si así lo acuerdan por lo menos cuatro de los cinco miembros del Comité, se podrá conceder una aprobación con distinción".

Se aprueba con Distinción

Observaciones:

Debe incorporar las observaciones realizadas por el Tribunal

Original: Estudiantes, copia: Escuela de Economía

Resumen Ejecutivo

Este estudio realiza, mediante un modelo logit multinivel multinomial y uno binomial, un análisis de la relación de probabilidad que las variables de contexto socio-demográfico, institucional y geográfico tienen con el rendimiento académico de los estudiantes en el primer curso de matemática universitaria en la Universidad de Costa Rica, seleccionando las carreras con alto contenido matemático, con datos de la Universidad de Costa Rica y el Ministerio de Educación. La regresión se efectuó en dos niveles, el del estudiante y el del colegio, diferenciando entre colegios cercanos a la sede Rodrigo Facio y para colegios lejanos a ésta, con el fin de evaluar desigualdades entre tales colegios. Se encontró que los efectos de las variables personales, como el sexo y el capital cultural, y sociales, como la pobreza, tienen un efecto directo y significativo sobre el desempeño académico, mientras que las variables institucionales inciden de manera desigual según la ubicación geográfica del estudiante, con la excepción de la variable MATEM, que resultó significativa en todos los modelos. Por otro lado, no se encontraron relaciones significativas de las variables geográficas incluidas, mientras que se constató que se tienen diferencias significativas en el rendimiento de estudiantes provenientes de un mismo colegio, validando el uso de modelos multinivel.

Palabras clave: Modelos multinivel, Modelo multinomial ordenado, efectos geográficos, efectos institucionales

Abstract

This investigation analyzes, through multinomial and binomial multilevel logit models, the likelihood relationship of students' socio-demographic, institutional and geographic variables on the academic attainment in their first mathematic class at the Universidad de Costa Rica. Data was provided by the Universidad de Costa Rica and the Ministerio de Educación, selecting only students in careers heavily loaded with mathematic courses. The econometric regression was done in two levels, the pupil one and the school one, differentiating for schools located near the Rodrigo Facio campus and for schools far away from it, in order to explore inequalities between those schools. It was found that the effects of personal variables, like gender and cultural capital, and social ones, like poverty, are direct and significant on the academic performance of the students, while institutional variables have an unequal effect on attainment depending on the geographical location of the student, with the exception of the MATEM variable, which ended up being significant among all models. On the other hand, there were no significant relationships found from the geographical characteristics included, while it was verified that there are significant differences on college attainment among students from the same school, thus validating the use of multilevel models.

Key Words: Multilevel models, ordered multinomial models, geographic effects, institutional effects

Derechos de Propiedad Intelectual

Los resultados obtenidos en esta obra son propiedad exclusiva del autor. Queda prohibida la reproducción parcial o total de esta obra bajo cualquier medio, sin la autorización expresa del autor, bajo pena de incurrir en los derechos de propiedad intelectual.

Se concede a la Universidad de Costa Rica y a cualquier dependencia de la Universidad, el derecho no exclusivo de utilizar esta obra para propósitos propios de su actividad académica.

Dedicatoria

Dedico el presente trabajo de a mi madre, que en paz descansa, quien con sus esfuerzos y consejos me convirtió en el profesional que soy. Sé que le habría gustado ver la culminación de esta etapa, y aprovecho para rescatar su (inadvertido) aporte a esta investigación, al siempre repetirme “no es el colegio el que hace triunfar al estudiante, es el estudiante el que por sí mismo, si se lo propone, triunfa”.

Marcos David Castillo Bastos

Índice General

Introducción	1
I. Antecedentes	5
II. Metodología	9
a. Modelo binomial multinivel	11
b. Modelo multinomial multinivel.....	12
c. Procedimientos de Estimación.....	13
d. Datos	13
e. Modelaje de los datos.....	24
III. Resultados	25
a. Modelo Binomial Multinivel.....	27
b. Modelos Multinomial Ordenado Multinivel	31
Conclusiones	38
Referencias.....	42
Anexos.....	45
Anexo 1: Distribución de Frecuencias de las variables dependientes de los modelos.....	45
Anexo 2: Matriz de Correlaciones de las variables del Modelo	46
Anexo 3: Mapa por colegio de procedencia, según cantidad de estudiantes en la base de datos ...	47
Anexo 4: Análisis de Variancia Colegio Lejano en Nota Ordinaria	48
Anexo 5: Mapa de colegios de procedencia por cantidad de estudiantes con nota categoría “Alta”	49

Introducción

Esta investigación tiene como fondo temático el rendimiento académico en el primer curso de Matemática en la Universidad de Costa Rica (UCR), buscando explicarlo bajo la perspectiva del contexto socio-demográfico, institucional y geográfico del que provenían los estudiantes antes de llevar el curso.

El marco teórico bajo el que se realiza esta investigación es la teoría del capital humano. Esta teoría se encuentra bajo los supuestos de existencia de un mercado de trabajo y de capitales perfectamente competitivos. Theodore Schultz es reconocido como el primer autor en señalar el capital humano como sinónimo de educación y formación de las personas, él explica que los seres humanos, al invertir en sí mismos, aumentan las posibilidades de mejorar su bienestar (Schultz, 1961).

La teoría del capital humano expone la importancia que tiene la educación en los diferenciales de ingreso salarial de los trabajadores, pues trabajadores más calificados reciben un salario real más alto, al tener una productividad marginal laboral mayor (Lucas, 1988). Asimismo esta mayor retribución se debe a que los individuos incurren en diversos gastos de educación y a que experimentan un alto costo de oportunidad por permanecer económicamente inactivos cuando están estudiando, a la espera de que su formación les brinde salarios más elevados en el futuro (Becker, 1964). Bajo esta visión, resulta directo entender que en la coyuntura actual, las personas busquen educarse con el fin de obtener trabajos con altos salarios en el mercado laboral.

Asimismo, diversos organismos internacionales, tales como el Banco Mundial y la Organización para la Cooperación y el Desarrollo Económico, han establecido que en los países en vías de desarrollo la educación debe buscar satisfacer la demanda de trabajadores capacitados, al garantizar que las personas sean capaces de realizar las acciones que el mercado laboral les requerirá (Vásquez Olivera, 2015).

Una población que ha dedicado más tiempo a educarse, cuenta con más capital humano, con lo que logran tener acceso a puestos de trabajo más calificados y su bienestar aumenta. De aquí se tiene la importancia de que el rendimiento académico de la población en la universidad sea alto, con el fin de contar con una fuerza laboral educada y apta para trabajos que exigen un mayor nivel de conocimiento. Un ejemplo en el caso costarricense es el de la compañía INTEL, la cual pasó de requerir empleados con educación universitaria y empleados con formación técnica, a demandar empleados con educación universitaria en su mayoría (Cordero, 2016) . Este cambio en la demanda por trabajadores calificados conlleva a un aumento en sus salarios y en el bienestar de sus empleados, pero también implica una presión a que la población costarricense se eduque profesionalmente.

En Costa Rica, son las carreras científicas en su mayoría las que, en los últimos años, han resultado codiciadas entre los estudiantes por brindar acceso a puestos con salarios altos (Agüero & Castillo, 2015). Lo que resulta relevante para esta investigación, es que estas carreras científicas presentan la condición de tener una alta carga de cursos de matemática en su plan de estudios, por lo que la buena calidad de esta asignatura toma mayor relevancia.

Entrando más a fondo a la realidad del país, históricamente el rendimiento en los cursos de matemática ha sido de los más bajos en la Universidad de Costa Rica y demás universidades públicas del país (Alpízar & Alfaro, 2013); lo que representa, dado lo explicado previamente, un reto para las instituciones de educación del país. Las escuelas de matemática de las universidades públicas han realizado avances, proyectos e innovaciones para mejorar este rendimiento (Vicerrectoría de Docencia, 2009); sin embargo, las investigaciones en este campo han sido reducidas, por lo que resulta de importancia analizar los factores que afectan en el rendimiento académico en estos cursos, y a partir de este análisis, diseñar proyectos o políticas que ayuden a mejorarlo.

Resulta también de interés el hecho de que el rendimiento bajo en matemática se presenta desde el nivel colegial en el país. Resultados recientes en las pruebas del Programa Internacional de Evaluación de Estudiantes (PISA) han mostrado que el país se aleja de los estándares definidos por la Organización para la Cooperación y el Desarrollo Económicos

(OCDE) en la prueba de matemática, lo que ha presionado al Ministerio de Educación Pública del país a realizar reformas con el fin de mejorarlo (Fernández E. , 2013).

Para especificar “rendimiento académico”, se encuentra variedad de definiciones: por ejemplo, rankings de estudiantes, duración para terminar una carrera o, las calificaciones obtenidas en los cursos. Las calificaciones son, usualmente, la variable utilizada para aproximar el rendimiento académico, pues la calidad educativa del individuo en una asignatura es “medida” mediante instrumentos estandarizados (pruebas) que evalúan si el conocimiento del estudiante en la disciplina es el esperado para los estándares establecidos, por lo que las calificaciones representan el resultado dado por el estudiante en los centros de enseñanza (Contreras, Gallegos, & Meneses, 2009) (Martínez, 2002) (Friedman, 1961). Para efectos de esta investigación, el rendimiento será aproximado mediante la calificación final obtenida en el primer curso universitario de matemática.

Al pensar en este tipo de rendimiento académico surgen dudas de si existen diferencias significativas que influyan en el rendimiento en matemática entre los estudiantes que ingresan a este tipo de carreras, así como sobre la importancia que tiene el entorno social del estudiante en el rendimiento académico.

Otra área de interés es cómo inciden las diferencias geográficas de los estudiantes en su rendimiento académico, o aún más, si el efecto de las variables de contexto varía según la ubicación geográfica del estudiante. Este tema es de relevancia porque diversos estudios recientes han notado el efecto que la distancia y la localización de los estudiantes genera sobre el desempeño educativo (Öckert, 2012). La Universidad de Costa Rica se caracteriza además por recibir estudiantes de múltiples zonas del país, por lo que la realidad geográfica varía entre estudiantes.

A raíz de estas inquietudes, se formula el objetivo general de la investigación: Analizar la relación socio-demográfica y espacial entre el rendimiento de los estudiantes en los cursos de matemática de las carreras que requieren una alta carga matemática y el colegio de procedencia.

Para alcanzar este objetivo, se estipularon dos objetivos específicos, en los cuales se orienta este trabajo:

1. Examinar la relación de las variables de contexto demográfico, institucional, social y geográfico en el rendimiento académico de los estudiantes
2. Determinar si existen diferencias significativas de las relaciones del contexto en el rendimiento según la ubicación espacial del estudiante antes de entrar a la Universidad

A partir de estos objetivos y según la revisión de literatura realizada, se propusieron las siguientes hipótesis de investigación:

- i. Estudiantes con características de procedencia distintas, presentan el mismo rendimiento académico
- ii. Estudiantes provenientes de un mismo colegio tienen el mismo rendimiento académico
- iii. Las variables del contexto colegial (características del colegio de procedencia) mantienen su significancia en el rendimiento académico sin importar la ubicación geográfica del colegio

Para llevar a cabo la comprobación de las hipótesis, se establecen relaciones entre las variables y la probabilidad de aprobar el primer curso de matemática, sin involucrar relaciones de causalidad. El análisis solo se realizó en carreras con alta carga matemática en primera instancia por estar compuestas por estudiantes que tendrían una afinidad más homogénea por las matemáticas; y en segunda, ya se ven presionados a aprobar el primer curso de matemática, al ser requisito de cursos posteriores. En la sección de datos se detallará más sobre esta selección de carreras.

Las variables dependientes en los modelos son de tipo discretas, por lo que se emplearon modelos logit: uno binomial, donde la variable dependiente es la condición de aprobar el curso de matemática, y uno multinomial ordenado, donde se crearon rangos de calificaciones.

Se utilizaron modelos econométricos multinivel, los cual brindan una serie de ventajas para la investigación en economía educativa, en especial por tomar en cuenta la estructura jerárquica de las instituciones educativas, y por permitir realizar análisis más desagregados y profundos del efecto de cada variable en el rendimiento educativo. (Dickey & Houston, 2010)

Además, se aplicó una clasificación de los modelos según la ubicación geográfica del colegio de procedencia, diferenciando entre colegios cercanos y lejanos a la sede Rodrigo Facio de la Universidad de Costa Rica, con el fin de determinar si se rechaza o no la tercera hipótesis de investigación.

I. Antecedentes

Como se verá a continuación, diversos estudios han analizado el efecto del contexto del estudiante en su rendimiento universitario; sin embargo, pocos han tomado en cuenta las variables geográficas y espaciales en sus análisis, especialmente en la región latinoamericana.

El estudio de Vargas (2010), realizado en la Universidad de Costa Rica, respecto al rendimiento universitario en matemáticas en la Universidad Nacional de Ingeniería de Nicaragua, obtuvo que los factores de mayor relevancia en el rendimiento académico fueron los promedios anteriores del estudiante, como los de bachillerato, las notas de admisión y el promedio ponderado en el colegio. Factores institucionales como la razón alumno/profesor en el curso o la edad del profesor impartiendo el curso mostraron un efecto inverso en el rendimiento del estudiante en los cursos de matemáticas. Otro factor de importancia fue la modalidad del nombramiento del profesor, evidenciando que los profesores en condición permanente (en propiedad) suelen tener estudiantes con mejor rendimiento.

En el caso costarricense, se han realizado estudios sobre los factores asociados al rendimiento académico en matemática en la educación general básica, incluyendo dimensiones personales, familiares y de la escuela o colegio, como el estudio para el Estado de la Educación de Oviedo (2012), que encontró efectos significativos en el rendimiento según la dependencia del centro educativo (público o privado), así como el agrado a la asignatura y la escala de satisfacción del

director. No obstante, a nivel universitario, el número de estudios se reduce y se pierde especificidad en los análisis.

El estudio de mayor comparabilidad a nivel nacional con la investigación presente es el de Moreira (2013), el cual realiza un análisis multinivel de los factores de contexto, admisión y procesos asociados en el rendimiento en los cursos de matemática del Instituto Tecnológico de Costa Rica (ITCR). Este estudio encuentra diferencias en el rendimiento entre estudiantes según dependencia económica del colegio de procedencia (público o privado), por nota de admisión (desagregada por el componente de Matemática) del ITCR y por edad del docente que imparte el curso de matemática.

Para la Universidad de Costa Rica (UCR) se encontró en el estudio de Garbanzo (2014) que la modalidad del colegio de procedencia, el nivel socioeconómico y la nota de admisión tenían correlaciones significativas con el rendimiento académico general de los estudiantes universitarios. Además, factores del ambiente familiar también mostraron una correlación fuerte, tales como la motivación y las conversaciones académicas en la familia.

En otro tipo de investigaciones se ha analizado cómo la transición entre la secundaria y la educación universitaria puede incidir en el rendimiento del estudiante, especialmente en el primer año de carrera. Para el caso costarricense, Mora (1998) mostró que la transición no es un proceso integral como el de la escuela al colegio, sino que no parece haber un enlace entre ambos, lo que afecta el desempeño de los estudiantes, en especial de los menos maduros.

De manera similar se ha estudiado cómo el llevar cursos de matemática en el colegio aumenta la probabilidad de completar la carrera, al brindarle al estudiante mayores habilidades iniciales (Trusty & Niles, 2003). En Costa Rica, el proyecto MATEM que la Escuela de Matemática inició en 1985 con el fin de preparar a los estudiantes de IV ciclo en el área de matemáticas para los cursos universitarios, ha mostrado múltiples beneficios para estos estudiantes (Alpizar & Alfaro, 2013). Sin embargo, no todos los estudiantes del país tienen acceso a este programa.

Valverde (2005) encontró que el capital cultural del estudiante (definido a partir del tipo de colegio de procedencia, la educación de los padres y el tipo de conversaciones en casa) tenía

un efecto significativo y moderado en el rendimiento académico en estudiantes de la UCR. Además halló que el sexo también resultaba ser un factor determinante. De igual forma los factores institucionales, como desempeño académico del profesor y la razón alumnos/profesor, resultaron significativos. El estudio mostró que el desempeño no depende solamente de la capacidad intelectual de los estudiantes, sino también de su capital cultural y su aptitud (medido por los resultados de la prueba de admisión a la UCR).

Ramírez (1993) en un estudio aplicado en las carreras de derecho, farmacia, medicina e ingeniería industrial, utilizó el tiempo requerido para terminar una carrera como definición de rendimiento académico, y reveló que los estudiantes provenientes de colegios privados presentaban mejor rendimiento en la mayoría de carreras.

Como se mencionó anteriormente, para la región, los estudios que muestran el efecto de las variables geográficas y espaciales en el rendimiento académico son muy limitados; sin embargo, se hallaron estudios importantes fuera de la región latinoamericana. Respecto a este factor geográfico, se encuentran dos temas de atención: primero, el del entorno del estudiante (lugar donde vive), y segundo, la distancia respecto a la universidad.

El primer tema se refiere al efecto del ambiente en el que vive el estudiante. Existen estudios de la relación geográfica-demográfica en la decisión de ingresar a la universidad (Rephann, 1996), que muestra que en zonas lejanas a centros urbanos, la probabilidad de decidir entrar a una universidad es baja; así como los efectos de vivir en áreas urbanas o rurales en el rendimiento de los estudiantes de secundaria (Cheers, 1990). No obstante, es necesario estudiar estos efectos en el rendimiento a nivel universitario en matemática.

Por otro lado, un estudio realizado en la UCR, aplicado a la Universidad Autónoma de Guerrero, México, utilizando un modelo multinivel, mostró que las variables socio-demográficas de los estudiantes (incluyendo sexo, lugar de procedencia y educación de los padres) no tenían un efecto significativo en el rendimiento académico general de los estudiantes, sino que eran más bien las variables psicológicas, pedagógicas e institucionales las que más influían. El estudio indagó también acerca del efecto del desplazamiento de los estudiantes universitarios a la hora de entrar a la universidad (si dejaban de vivir en casa de

sus padres por la distancia con el centro) sobre el rendimiento, no se hallaron resultados concluyentes (Barragán, 2008).

Respecto al segundo tema, la literatura es aún más escasa. Los efectos que tiene la distancia entre la residencia del estudiante y su universidad sobre el rendimiento no han sido estudiados a profundidad, a pesar de que el desplazamiento en el que incurren los estudiantes, ya sea para ir a clases o a estudiar, tiene una gran importancia pues absorbe tiempo que puede ser usado para invertir en educación.

Un estudio realizado en Alabama, Estados Unidos, encontró que características agregadas del estudiante como el colegio de procedencia y la ubicación afectan tanto la decisión de ingresar a la universidad, como la de persistir y obtener un título. Se mostraron relaciones entre las características a nivel agregado de los colegios de procedencia (como el rendimiento promedio del colegio y la proporción de estudiantes que desea entrar a universidades públicas) y las probabilidades de entrar, persistir y graduarse en una universidad. El estudio mostró que factores individuales como el sexo (las mujeres tienen una mayor probabilidad de éxito) y el ingreso familiar (a mayor ingreso mayor probabilidad de éxito) son significativos para explicar el rendimiento. Esta investigación también incluyó como variable la cercanía del colegio de procedencia respecto a alguna universidad, y ésta resultó tener un efecto significativo y negativo en la probabilidad de éxito (a mayor distancia menor probabilidad) (Johnson, 2008).

Otros estudios analizan los efectos de distancia en la decisión de asistir a la universidad como el de Öckert (2012), y de continuar en ella, como el de López (2009), pero en cuanto a rendimiento solo se analiza la persistencia del estudiante. El trabajo de Falch, Lujala y Strom (2011) es el que analiza con más detalle el efecto de las diferencias geográficas sobre el logro educativo en el colegio. Para medir las distancias entre los estudiantes y el centro educativo utilizan la red de caminos para unir los puntos y luego utilizan datos de tiempos de transporte para llegar de un lugar a otro, en lugar de solo la distancia, y encuentran que menores tiempos de viaje tienen un efecto positivo en la probabilidad de graduación.

II. Metodología

Como se señaló, el objetivo de este trabajo es determinar si variables del contexto institucional y demográfico tienen relación con el rendimiento académico universitario en matemática. Para el modelaje econométrico se deben tomar en cuenta dos temas: primero, la manera en que se va a definir y tratar el rendimiento académico como variable dependiente; y segundo, la estructura jerárquica de la educación costarricense (estudiante – colegio).

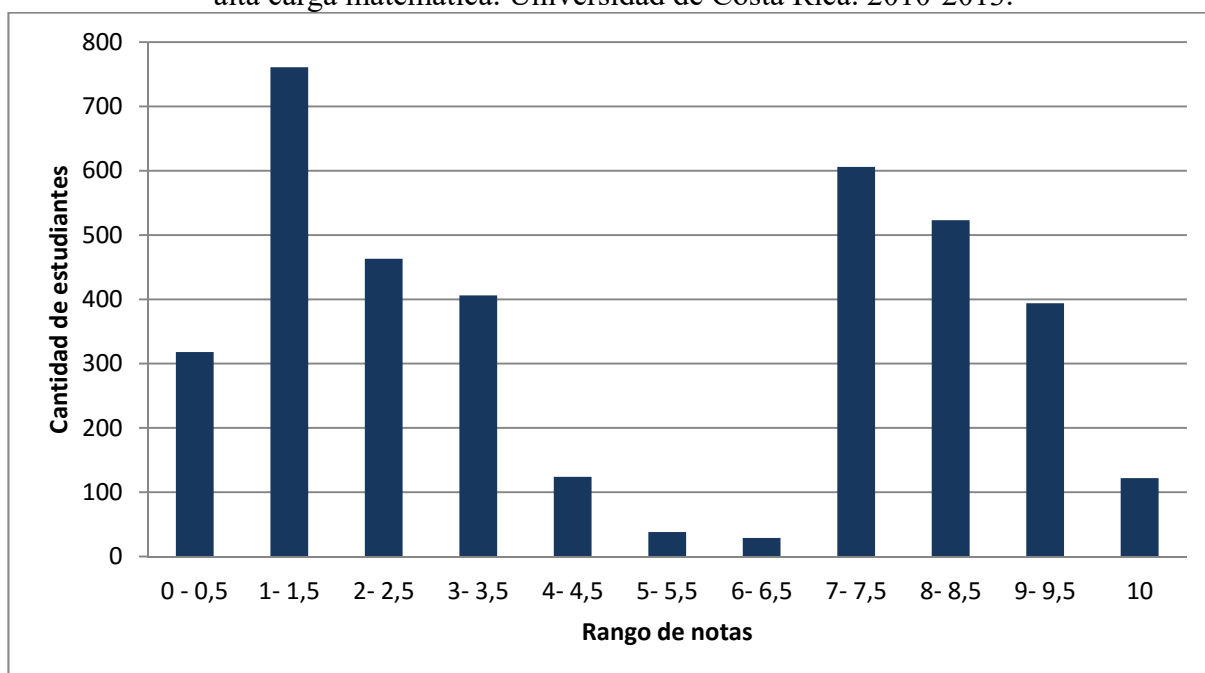
Para el primer punto, como se mencionó previamente, y siguiendo a Vargas (2010), Moreira (2013) y Garbanzo (2007) entre otros, el rendimiento académico es aproximado por las calificaciones obtenidas, en específico las de los cursos de matemáticas, al considerarse un resultado objetivo y significativo del desempeño del estudiante en la asignatura.

Debido a que las calificaciones de los cursos en la Universidad de Costa Rica son discretas (de “0” a “10” redondeadas a intervalos de 0,5), y en especial porque su distribución de frecuencias no tiende a una normal, sería inapropiado utilizar un modelo lineal. El Gráfico 1 muestra la distribución de frecuencias de las calificaciones en el primer curso de matemática. Como se puede apreciar en el histograma, la distribución de notas no tiende a una normal, pues se acumula en notas muy bajas (menores a 3) y en notas de 7, mientras que entre 5 y 6 el número de datos es relativamente reducido. Una explicación para este comportamiento es la perseverancia del estudiante, y la opción de ampliación, ya que estudiantes que obtuvieron notas bajas en las primeras evaluaciones de los cursos podrían desistir de llevar el curso (por lo que la nota final resultaría muy baja), mientras que otros (menos) estudiantes perseverarían y podrían tener la opción de ampliación¹, con lo que en caso de aprobarla, obtendrían un 7 de nota final (lo que reduce aún más la frecuencia de nota “6”). Este fenómeno, sin embargo, requiere de un estudio más profundo para explorar sus causas con mayor detalle.

¹ La opción de ampliación es un derecho brindado a los estudiantes que obtuvieron notas de entre 6 y 6.5 en la nota final, con lo que pueden realizar una prueba que, en caso de ser aprobada con 7 o más, le permitirá subir su nota final a 7 y aprobar el curso. (Universidad de Costa Rica, 2001)

Es por esta razón que se opta por dos modelos logit: uno usando como variable dependiente una categoría *dummy* de si el estudiante aprobó el curso o no (binomial) y otro usando categorías generadas a partir de los rangos derivados de la distribución de frecuencias de las notas obtenidas por los estudiantes de la UCR (multinomial ordenado). Más adelante se detallará la manera en que se crearon los rangos.

Gráfico 1: Histograma de Nota Ordinaria en el primer curso de matemática en carreras con alta carga matemática. Universidad de Costa Rica. 2010-2013.



Fuente: Elaboración propia

Respecto al segundo punto, la respuesta viene dada por los modelos multinivel. Este tipo de modelos es utilizado frecuentemente en el área de economía de la educación, pues resuelve la complejidad impuesta por la estructura organizacional del sistema educativo, al diferenciar los distintos niveles o conglomerados de los estudiantes. En un colegio, cada estudiante posee características propias que lo diferencian de sus compañeros, pero a su vez, comparte relaciones en su colegio (su conglomerado), por lo que ciertas variables institucionales, sociales y geográficas se encuentran correlacionadas entre los individuos, y sería incorrecto

tratar todas las variables como independientes cuando se sabe que hay relaciones entre ellas (Barragán, 2008). De esta manera, los modelos a estimar son los de multinivel binomial y multinivel multinomial ordenado, que se definen a continuación.

a. Modelo binomial multinivel

Para la explicación metodológica de los modelos, se utilizaron como referencia *Multilevel Statistical Models* (2003) y *A User's Guide to MLWin* (2015).

El modelo binomial tiene como variable dependiente la variable “Aprueba” (y_{ij}), la cual es dicotómica y toma el valor de 1 si el estudiante “i” proveniente del colegio “j” aprobó su primer curso de matemática la primera vez que lo llevó, y 0 en cualquier otro caso.

La probabilidad de que un estudiante apruebe el curso viene dada por

$$\pi_{ij} = \Pr(y_{ij} = 1)$$

De esta forma el modelo a estimar sigue la forma

$$\text{logit}(\pi_{ij}) = \beta_{0j} + \beta_1 X_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

Así en un modelo de dos niveles, X representa el vector de variables explicativas, y el intercepto consiste de dos términos: el componente fijo β_0 y el efecto aleatorio u_{0j} , que es un componente específico para el segundo nivel (colegio). La diferencia de los modelos multinivel con los de un solo nivel, es la existencia del efecto aleatorio en los colegios. Este efecto aleatorio permite que dentro de un mismo conglomerado (el colegio), las observaciones puedan variar en sus resultados.

El modelo asume una distribución normal de los residuos u_{0j} con media 0 y varianza $\sigma_{u_0}^2$.

b. Modelo multinomial multinivel

Un modelo multinomial ordenado se basa en la probabilidad *acumulada* de las respuestas, al contrario del modelo multinomial simple que trata cada categoría por separado. Las probabilidades acumulativas de respuestas se definen como:

$$E(y_i^s) = \gamma_i^s = \sum_{h=1}^s \pi_i^h; \quad s = 1, \dots, t - 1$$

Donde “s” corresponde a las categorías de la variable dependiente ($s = 1, \dots, t$) y y_i^s son las proporciones acumuladas observadas para el i-ésimo estudiante. Las probabilidades acumuladas se pueden expresar como:

$$\pi_i^h = \gamma_i^h - \gamma_i^{h-1}; \quad 1 < h < t$$

$$\pi_i^1 = \gamma_i^1; \quad \gamma_i^t = 1$$

El modelo comúnmente utilizado es el de *probabilidades proporcionales logit*, en el que:

$$\text{logit}(\gamma_i^s) = \alpha^s + (X\beta)_i$$

El cual implica que aumentos en el valor del componente lineal están asociados con probabilidades crecientes cuando s aumenta. Como el modelo a regresar es un modelo multinomial multinivel (de dos niveles), la especificación viene dada entonces por:

$$\text{logit}(\gamma_i^s) = \alpha^s + (X\beta)_{ij} + Z_{ij}u_j$$

Donde “j” representa el grupo (colegio) en el cual los individuos “i” se relacionan. De esta forma las probabilidades se van a poder expresar como

$$\pi_{ij}^h = \gamma_{ij}^h - \gamma_{ij}^{h-1}; \quad 1 < h < t$$

$$\pi_{ij}^1 = \gamma_{ij}^1; \quad \gamma_{ij}^t = 1$$

Para el análisis de los modelos es necesario indicar una categoría de referencia de la variable dependiente, ya sea la más baja o lo más alta. En este caso se eligió la categoría más alta como la de referencia, por lo que la interpretación de los datos se realiza en comparación con esta categoría. Para mayor información respecto este tipo de modelos consultar Goldstein (2003).

c. Procedimientos de Estimación

En los modelos multinivel de respuesta discreta, se utilizan métodos de cuasi-verosimilitud para su estimación, pues estos linealizan la respuesta discreta en una respuesta continua mediante una expansión de series de Taylor. Para realizar la linealización las aproximaciones pueden ser mediante el método de cuasi-verosimilitud marginal (MQL) o el de cuasi-verosimilitud predictivo (PQL), ya sean de primer o segundo orden de la serie de Taylor.

Para efectos de esta investigación, se utilizó la aproximación de mayor predicción, PQL de segundo orden, pues el procedimiento MQL ofrece la aproximación más cruda y usualmente se obtienen sesgos en los estimadores. (Rasbash, Steele, Browne, & Goldstein, 2015)

d. Datos

Los datos para la investigación fueron facilitados por la Escuela de Matemática de la UCR para los años 2003 a 2013, y de Admisión para los años 2010 a 2013. De ahí que, dado el cruce necesario para realizar el estudio, se analizó el periodo del 2010 al 2013. Estos datos incluyen la nota obtenida por cada estudiante en cada curso de matemática llevado, sexo, nota en el examen de admisión, colegio de procedencia y si llevó MATEM en el colegio; asimismo, incluye datos de si el estudiante realizó el examen de Diagnóstico de Matemática (DIMA), y en caso afirmativo, la nota obtenida en tal prueba.

Por otra parte, para los datos de los colegios se utilizó la Megabase de Secundaria del Ministerio de Educación Pública para los años 2009-2012 (se tomó un año de rezago en los datos de colegio, asumiendo que los estudiantes ingresaron a la universidad al año siguiente de

haberse graduado del colegio). Estos datos se encuentran disponibles en línea para el Cuarto Informe Estado de la Educación, que incluye la georreferenciación de los colegios realizada por el Programa de Investigación en Desarrollo Urbano Sostenible de la Universidad de Costa Rica (ProDUS-UCR) en el 2013. La Megabase de Secundaria contiene datos sobre infraestructura, matrícula, deserción, repitencia, aprobación, dependencia económica, asignaturas, contacto, entre otras características de todos los colegios del país, para cada año.

Además, se utilizaron los datos de pobreza por distrito del Censo 2011 obtenidos mediante la plataforma en línea del Instituto Nacional de Estadística y Censos (INEC).

Los criterios de inclusión de los estudiantes, cursos y carreras analizadas se describen a continuación. Primero, los estudiantes deben estar matriculados en carrera con alto contenido de matemáticas en su plan de estudios (al menos 4 cursos de la Escuela de Matemática), pues usualmente en estas carreras los cursos de matemáticas son requisitos de otros cursos del programa y se colocan en los primeros semestres del plan de estudios. Esta selección es relevante porque en estas carreras los estudiantes se encuentran presionados a aprobar tales cursos para no retrasar su ingreso al mercado laboral.

Segundo, se decidió sólo incluir las calificaciones obtenidas en el primer curso de matemática por primera vez, pues es cuando la experiencia colegial se encuentra más reciente y no se tiene el sesgo de que el estudiante haya repetido el curso o que haya desistido de llevar la carrera (con lo que solo se incluirían los estudiantes exitosos).

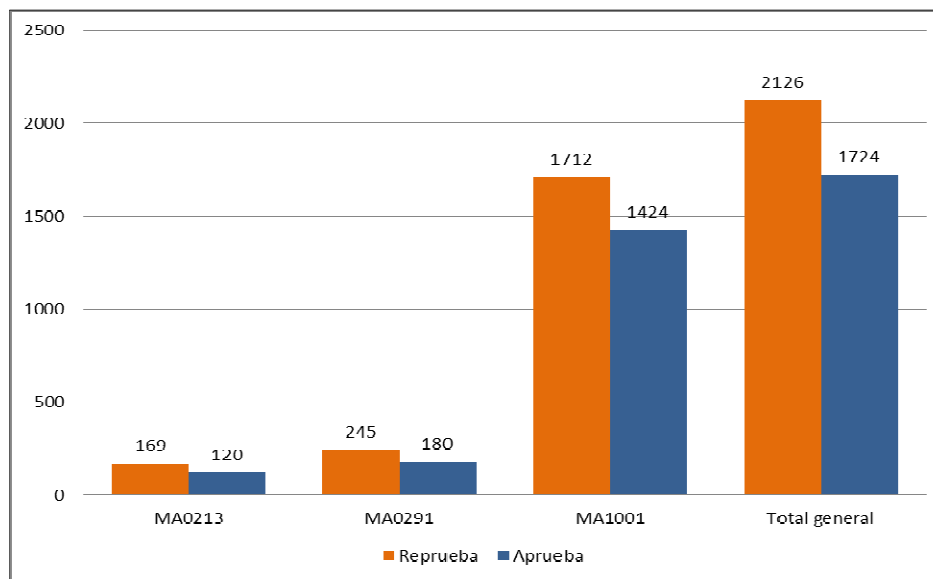
Tercero, sólo se incluyeron los estudiantes matriculados en la sede Rodrigo Facio en San Pedro de Montes de Oca, la sede central de la UCR. Esto obedece a dos motivos relacionados con la disponibilidad de información: primero, en las sedes regionales son pocas las carreras con al menos 4 cursos de matemática en el plan de estudios, y segundo, incluir a otras sedes hubiera dificultado considerablemente el análisis geográfico.

Finalmente, no se incluyeron estudiantes de colegios del extranjero, que no indicaron su colegio de procedencia o que para dicho colegio no se contara con datos en la Megabase; o

aquellos estudiantes que convalidaron el primer curso, lo llevaron por suficiencia o lo retiraron formalmente.

Las carreras seleccionadas fueron: economía (modalidades “bachillerato” y “bachillerato y licenciatura”²), física, ingeniería topográfica, ingeniería agrícola, ingeniería eléctrica, ingeniería mecánica, ingeniería civil, ingeniería industrial, ingeniería química y ciencias de la computación e informática. Para estas carreras se seleccionó el primer curso de matemática según los planes de estudios para cada caso, resultando seleccionados los siguientes cursos: Cálculo 1 (MA1001), Matemática para Economía y Estadística (MA0213) e Introducción a la matemática para Computación (MA0291)³. El gráfico 2 muestra la cantidad de estudiantes por cada carrera, según su condición de aprobación.

Gráfico 2. Cantidad de estudiantes por condición de aprobación, según curso de Matemática, Universidad de Costa Rica. 2010-2013



Fuente: Elaboración propia

² A partir del año 2010 la carrera pasó de ser “Bachillerato y Licenciatura en Economía con varios énfasis” a solo “Bachillerato en Economía”.

³ Este curso de la carrera de Ciencias de la Computación e Informática puede ser convalidado por el curso MA0129; sin embargo, solo se tomó en cuenta el curso estipulado en el plan de estudios, pues es el que la mayoría de estudiantes matriculan.

Como se explicó anteriormente, la variable dependiente de interés no es continua, por lo que para el modelo multinomial fue necesario generar una nueva variable categórica agrupando las notas obtenidas. Esta agrupación se realizó tomando en cuenta las clasificaciones que la UCR adjudica a las notas de los cursos (Universidad de Costa Rica, 2001) y la distribución de frecuencias de las mismas. De esta forma, se crearon 5 grupos o categorías, que se muestran en el Cuadro 1.

Cuadro 1. Rangos para la variable dependiente “Nota Categórica” en el primer curso de matemática

Nombre de la Categoría	Rango de Nota Obtenida	Frecuencia
Cero	De 0	311
Bajo	De 0,5 a 2,0	1202
Insuficiente	De 2,5 a 6,5	600
Aceptable	De 7,0 a 7,5	841
Alto	De 8,0 a 10,0	768

Fuente: Elaboración propia

Es importante notar que la nota de “0” puede indicar estudiantes que dejaron el curso (sin realizar el proceso en Registro), la carrera o la universidad de forma informal. Si bien esto no reflejaría el rendimiento del estudiante en el curso, sí es un factor importante de estudio, por lo que fue incluido en el análisis. A su vez, se creó una variable dicotómica que solo indica si el estudiante aprobó o no el primer curso, denominada “Aprueba”.

El cuadro 2 presenta un resumen de las 8 variables elegidas como explicativas, más la dependiente, así como el nivel jerárquico en el modelo, descripción, el tipo de variable, el efecto esperado en el rendimiento (según la evidencia empírica y la teoría) y las observaciones al respecto. La justificación de la inclusión de cada variable se explica posteriormente. En el caso de las variables dicotómicas, se eligió como categoría de referencia la que fuera señalada por la literatura como la que afecta positivamente el rendimiento académico, con la excepción de las variables Pobreza y Colegio Lejano, pues para efectos del análisis e interpretación, resulta más sencillo valorar el efecto negativo de tales categorías en el rendimiento.

Cuadro 2. Variables incluidas en el modelo

Nivel	Variable	Descripción	Tipo	Efecto Esperado	Observaciones/ Fuente
Dependientes	Nota Categórica (<i>NotaCat</i>)	Nota obtenida en el primer curso de Matemática, agrupada en 5 grupos	Categórica (Categorías explicadas en el cuadro 1).		Se obtuvo a partir de la variable Nota Ordinaria de la base de datos de la Escuela de Matemática
	Aprueba (<i>Aprueba</i>)	Muestra si el estudiante aprobó o no el primer curso de Matemática	Dicotómica 0=Nota de 7 o menos 1=Nota de menos de 7		Se obtuvo a partir de la variable Nota Ordinaria de la base de datos de la Escuela de Matemática
Primer Nivel: Estudiante	Mujer (<i>Mujer</i>)	Indicador de si el estudiante es mujer	Dicotómica. 0=Hombre 1=Mujer	Positivo	Base de datos de Admisión
	Nota de Admisión (<i>NotaAdmi</i>)	Nota obtenida en el examen de admisión	Continua, acotada entre 442,29 y 800.	Positivo	Base de datos de Admisión.
Segundo Nivel: Colegio	MATEM (<i>MATEM</i>)	Indicador de si el estudiante llevó el curso MATEM en el colegio	Dicotómica. 0=No llevó MATEM 1=Llevó MATEM	Positivo	Base de datos de la Escuela de Matemática
	Colegio Privado (<i>CPrivado</i>)	Indica si el estudiante se graduó de colegio privado	Dicotómica: 0=Público 1=Privado	Positivo	Los estudiantes de colegio subvencionados fueron agrupados como privados, pues es así como lo realiza la Universidad de Costa Rica. Se obtuvo de la

Nivel	Variable	Descripción	Tipo	Efecto Esperado	Observaciones/ Fuente
					Base de datos de Admisión.
	Porcentaje de profesores titulados (<i>CTitulados</i>)	Muestra la razón de profesores titulados respecto al total de profesores en el colegio de procedencia	Continua acotada entre 0 y 1	Positivo	Se obtuvo de la Megabase y se unió para cada colegio según el año de graduación de cada estudiante
	Laboratorio de Informática (<i>CInformática</i>)	Indica si el estudiante asistió a un colegio que tenía laboratorio de informática	Dicotómica. 0=No tenía 1=Sí tenía	Positivo	Se obtuvo de la Megabase y se unió para cada colegio según el año de graduación de cada estudiante
	Colegio Lejano (<i>CLejano</i>)	Es una <i>proxy</i> que muestra si el estudiante se tuvo que mudar de residencia para estudiar en la Universidad, pues su colegio de procedencia está lejano a la sede Rodrigo Facio	Dicotómica. 0=No se desplazó (Colegio dentro del rango definido) 1=Se desplazó (Colegio fuera del rango)	Negativo	El colegio de procedencia se estableció según la base de colegios geo-referenciados, y se separó a los que estuvieran a más de 50km de la sede Rodrigo Facio ⁴

⁴ Como referencia, el Anexo 3 muestra un mapa con los colegios de procedencia por cantidad de estudiantes en la base de datos.

Nivel	Variable	Descripción	Tipo	Efecto Esperado	Observaciones/ Fuente
	Distancia (<i>Distancia</i>)	Distancia en kilómetros entre el colegio de procedencia y la sede Rodrigo Facio	Continua que va de 0km a 50km	Negativo	La distancia es en línea recta y solo incluye los colegios dentro del rango definido. Del colegio de procedencia se estableció según la base de colegios geo-referenciados y se calculó la distancia mediante el software ArcGIS.
	Pobreza (<i>Pobreza</i>)	Indica si el estudiante proviene de un colegio en un distrito con altos niveles de pobreza según carencias críticas	Dicotómica 0=Distrito en el 80% con más bajos niveles de pobreza 1=Distrito entre el 20% con más altos porcentajes de pobreza	Negativo	Censo 2011 mediante la plataforma virtual del INEC

Fuente: Elaboración propia

La variable de sexo (Mujer) fue incluida en el análisis pues estudios anteriores han mostrado diferencias significativas en el rendimiento entre hombres y mujeres, entre ellos Barragán (2008), Moreira Mora (2013) y Garbanzo (2014). Estos estudios señalan un mejor rendimiento en el caso de las mujeres, por lo que se eligió la variable Mujer como variable de interés. Además, las variables edad y sexo son las más utilizadas en estudios de economía de la educación porque gran parte del comportamiento demográfico de un individuo está condicionado a estas dos variables (Ortega, 2001). En la base de datos, de 3784 estudiantes incluidos, 1091 fueron mujeres.

La “Nota de Admisión” se incluye como aproximación del historial académico del estudiante. Este historial es de importancia pues cuando los estudiantes tienen confianza en su capacidad, aceptan desafíos y retos en sus tareas, y persisten en alcanzar el éxito, lo que afecta el rendimiento académico (García & Doménech, 1997). La nota de admisión promedio para los estudiantes en la base de datos fue de 630,04 puntos.

En esta misma línea, la variable “MATEM” fue incluida pues los estudiantes que llevan este curso se encuentran mejor preparados para los cursos de matemática universitaria (Trusty & Niles, 2003); por lo que se quiere probar si, estadísticamente, esta variable tiene efecto en el rendimiento final del estudiante.

A nivel nacional, estudios como el de Moreira Mora (2013) y Garbanzo (2007); y en el marco internacional como Martínez (2002), entre otros, han mostrado que el provenir de colegio público o privado se relaciona con factores socio-económicos, culturales y educativos del estudiante y su entorno, los cuales generan diferencias en el rendimiento académico. Esta variable resulta de interés también por los recientes cambios que ha realizado el Ministerio de Educación Pública del país con el fin de mejorar el rendimiento académico en matemática a nivel de secundaria, cerrando la brecha entre el rendimiento de estudiantes de colegio público y privado (Fernández & Del Valle, 2013).

Tanto la literatura internacional (Card & Krueger, 1992) como los resultados obtenidos para Costa Rica, concuerdan en que una mayor cantidad de profesores titulados ayuda a mejorar el rendimiento de los estudiantes (Programa Estado de la Nación, 2013), por lo que se decidió incluir el porcentaje de profesores titulados en el colegio de procedencia.

De igual forma, se considera que el contar con laboratorio de informática, además de ser un aproximado de buena infraestructura, fortalece el rendimiento académico del estudiante (Programa Estado de la Nación, 2013).

Tal y como se expuso anteriormente, esta investigación explora si el factor geográfico incide en el rendimiento entre los estudiantes. En primera instancia se quiere estudiar si hay diferencias en el rendimiento entre estudiantes que tuvieron que mudarse de residencia para

estudiar en la universidad, y aquellos que se mantuvieron en su hogar. En segunda instancia, se analiza si la distancia que tienen que recorrer para llegar al centro de estudios afecta el rendimiento. Para el primer punto se creó la variable “Colegio Lejano” y para el segundo la variable “Distancia”; ambas se explican a continuación.

La variable “Colegio Lejano” se obtuvo a partir de la distancia lineal del colegio de procedencia con la sede Rodrigo Facio de la UCR. Al realizar el análisis de frecuencia de las distancias de los colegios, se encontró que a partir de los 50km de distancia de la Sede Central, ocurre un descenso importante en la cantidad de colegios con estudiantes en la base de datos, lo que concuerda con la literatura de que a más de 40 km de distancia (en zonas sub-urbanas o rurales) muchos estudiantes optan por no asistir a la universidad debido al largo tiempo de viaje (López, 2009).

Siguiendo este hallazgo, se realiza el supuesto de que, a partir de los 50km de distancia, los estudiantes que deciden ingresar en la sede Rodrigo Facio se mudan a una residencia más cercana a esta sede, en lugar de viajar al centro de estudio. La variable toma el valor de “1” si el estudiante proviene de colegio lejano, y “0” en cualquier otro caso.

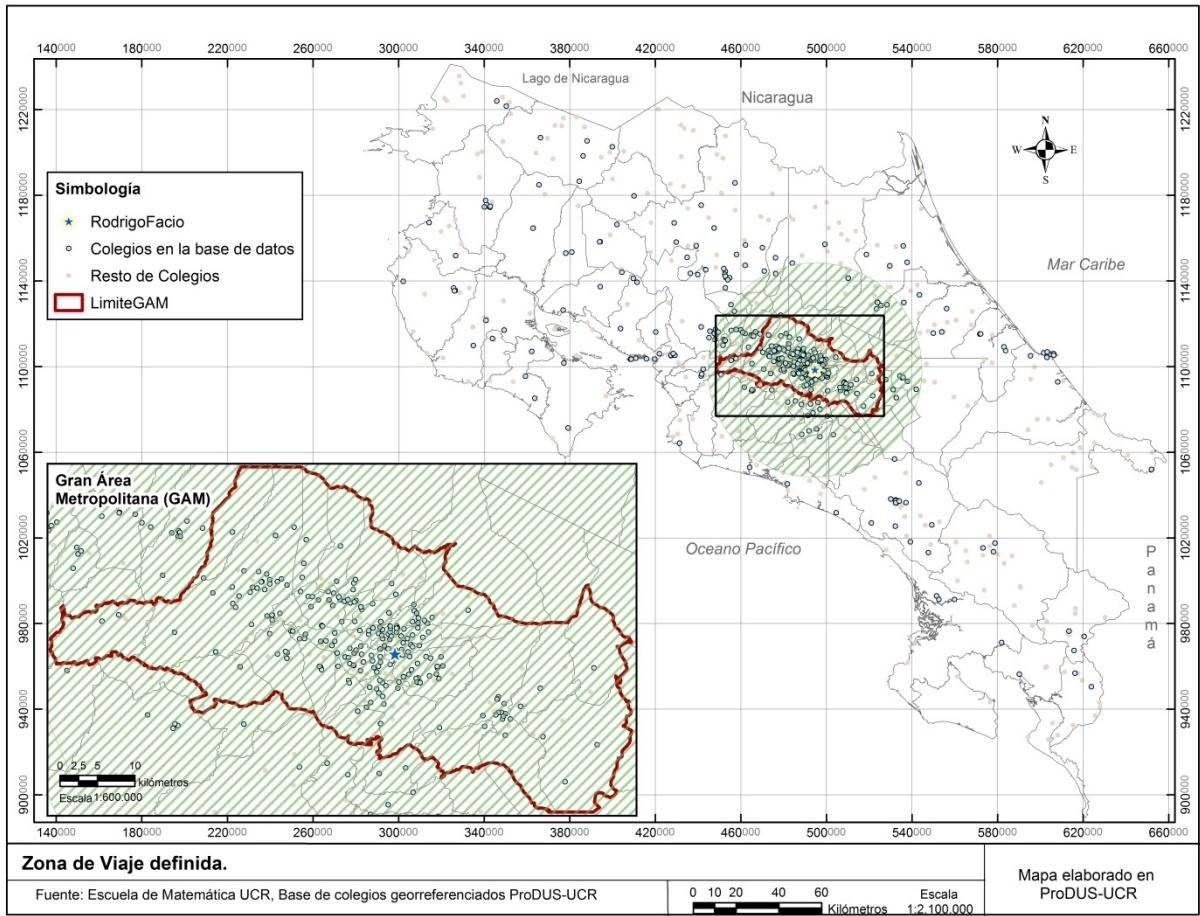
El Mapa 1 muestra los colegios de procedencia de los estudiantes incluidos en la base y el radio de 50km alrededor de la sede Rodrigo Facio como referencia.

Por otro lado, la variable “Distancia” se obtuvo de igual forma que la distancia lineal del colegio de procedencia respecto a la sede Rodrigo Facio; sin embargo, en este caso la variable es de tipo continua, y solo se utilizó para los modelos de colegio cercano, pues en el caso de estudiantes de colegio lejano resulta irrelevante.

Cabe resaltar que para ambas variables se está asumiendo que los estudiantes no se trasladaron grandes distancias para asistir al colegio, por lo que los colegios de procedencia son aproximadores (variables *proxy*) de las residencias de los estudiantes durante el colegio.

El cuadro 3 muestra la cantidad de colegios incluidos en la base de datos para cada una de las variables dicotómicas de segundo nivel de los modelos.

Mapa 1: Zona de viaje definida para el estudio



Cuadro 3: Cantidad de colegios según las variables dicotómicas definidas para los modelos

Variable	Valor de 0	Valor de 1	Total
MATEM	447	5	452
Colegio Privado	293	159	452
Lab. Informática	119	333	452
Zona de Pobreza	407	45	452
Colegio Lejano	248	184	*432

*No incluye los colegios no georreferenciados

Fuente: Elaboración propia

El cuadro 4 a continuación presenta los promedios para las variables continuas de segundo nivel de la base de datos.

Cuadro 4: Promedio de las variables continuas de segundo nivel

Variable	Promedio
Porcentaje de profesores titulados	92,20%
Distancia*	12,04km

*Solo para colegios cercanos

Fuente: Elaboración propia

Finalmente, los efectos socioeconómicos han sido mencionados reiteradamente en la literatura por su importancia en el rendimiento académico, ya que alumnos con dificultades sociales y económicas tienen más dificultad para estudiar y se rodean de un ambiente de bajo capital cultural y académico (Valverde, 2005).

La variable de pobreza incluida en el modelo, es una dicotómica a partir del ranking de la variable de pobreza según carencias críticas del INEC. En la dicotómica, el 20% de distritos con porcentajes de pobreza multidimensional más altos fueron los seleccionados con Pobreza=1, mientras que el restante 80% (así como todos los colegios privados) se codificó con Pobreza=0. El porcentaje de pobreza por carencias críticas, según lo definen Méndez y Trejos (2000), es una variable que representa el porcentaje de hogares con al menos una carencia, es decir, que no ha satisfecho una necesidad básica. Estas carencias incluyen cuatro dimensiones sociales: albergue (infraestructura de la vivienda), salud (acceso a higiene), educación (nivel de escolaridad o rezago escolar del hogar) y consumo (ingreso del hogar); por lo que este porcentaje resulta un indicador de la condición social del distrito en el que habitó el estudiante durante sus estudios de secundaria (Instituto Nacional de Estadística y Censos, 2013).

Si bien la variable de entorno social incluida resulta amplia y general (al ser por distrito y no por estudiante), es la mejor aproximación posible a la realidad socioeconómica del estudiante a nivel de datos, puesto que no se obtuvo información a nivel individual.

Finalmente, cabe destacar que al realizarse un modelo multinomial, y al encontrarse problemas de multicolinealidad de las variables Nota de Admisión y Porcentaje de Profesores Titulados⁵, estas variables fueron normalizadas, con lo que se facilitó la posterior interpretación de los resultados de los modelos. Similarmente, es importante notar que a pesar de que la base de datos de matemática incluye la variable DIMA, ésta no se incluyó en el análisis por falta de datos en las observaciones del año 2010. Además, para el período estudiado no era obligatorio realizar la prueba, por lo que una gran cantidad de estudiantes no la realizaron (35% de todas las observaciones).

e. Modelaje de los datos

En el caso específico de estudio, para los modelos multinivel se definieron dos niveles: el nivel del estudiante (i) y el del colegio (j). En el primero, el más básico y desagregado, cada característica es fija para cada estudiante. En este caso las variables del primer nivel (mostradas en el cuadro 2) son la Nota Categórica, Mujer y Nota de Admisión.

Respecto a la variable MATEM es necesario aclarar que si bien la variable se presenta a nivel del estudiante (nivel 1) en la base de datos, se encontró que en los colegios en los que se ofrecía la posibilidad de llevar MATEM, todos los estudiantes optaron por llevar el curso, con lo que no hubo variabilidad aleatoria dentro de los colegios. Por esta razón, se decidió no incluir MATEM como variable de nivel 1.

El segundo nivel, jerárquicamente más alto que el primero, es el del colegio de procedencia, en el cual se permite que las observaciones (estudiantes) tengan efectos aleatorios dentro del mismo colegio. De igual forma, las variables de segundo nivel de este caso, se encuentran en el cuadro 2 y son: Colegio Privado, Porcentaje de Profesores Titulados, Laboratorio de Informática, Colegio Lejano, Distancia y Pobreza. La variable Pobreza, al estar definida distritalmente, en teoría debería ser considerada de nivel 3; sin embargo, dada la cantidad de

⁵ Las proporciones de varianza de las variables Nota de Admisión y Porc. de Prof. Titulados fueron de 0,80 y 0,16 del índice de condición de colinealidad respectivamente, lo que revela alta multicolinealidad de tales variables

distritos en el país, cada distrito solo incluye una reducida cantidad de colegios (de 1 o 2 en su mayoría). Por esta razón se consideró la variable como nivel 2, asumiendo que los estudiantes de distintos colegios de un distrito se rodean de características similares de pobreza.

Se ejecutaron dos tipos de modelos para la investigación, uno multinomial ordenado y uno binomial. Además, dado que se exploraron las diferencias espaciales entre los estudiantes, para cada tipo de modelo se estimaron tres regresiones a partir de distintas bases de datos: una base completa, una acotada a los colegios cercanos a la universidad y una acotada a los colegios lejanos. Para definir estas bases se utilizó la variable dicotómica “Colegio Lejano”, donde se filtró “Colegio Lejano=0” para la regresión acotada a los colegios cercanos y “Colegio Lejano=1” para la de colegios lejanos.

En total la base de datos para los modelos completos incluye 3784 observaciones, 2 variables dependientes (una para cada tipo de modelo) y 8 variables independientes (incluye Colegio Lejano), mientras que la base de datos de colegio cercano contiene 2646 observaciones, 2 variables dependientes y 8 variables independientes (incluye Distancia), y la de colegio lejano incluye 1076 observaciones, 2 variables dependientes y 7 variables independientes.

III. Resultados

Mediante el software SPSS se obtuvieron las estadísticas descriptivas de las variables del modelo, las cuales se muestran en el Cuadro 4. Además, en el anexo 1 se presenta la distribución de frecuencias de las variables dependientes de cada modelo (NotaCat y Aprueba), y en el anexo 2 la matriz de correlaciones de las variables.

Cuadro 4. Estadísticas Descriptivas para las tres bases analizadas

	Base Completo (n=3784*)		Base CCercano (n=2646)		Base CLejano (n=1076)	
	Media	Desviación estándar	Media	Desviación estándar	Media	Desviación estándar
Nota Ordinaria	4,27	3,32	4,48	3,33	3,78	3,25
Aprueba	0,43	0,50	0,46	0,50	0,36	0,48
Mujer	0,29	0,45	0,29	0,45	0,29	0,45
Nota Admisión	630,04	62,58	637,07	60,93	612,91	63,01
MATEM	0,03	0,18	0,04	0,20	0,01	0,10
Colegio privado	0,49	0,50	0,58	0,49	0,25	0,43
Porcentaje de profesores titulados	0,95	0,14	0,95	0,14	0,96	0,09
Colegio con Lab de Informática	0,65	0,48	0,66	0,47	0,62	0,48
Colegio Lejano	0,29	0,45				
Distancia			10,15	8,44		
Pobreza	0,02	0,15	0,00	0,03	0,08	0,28

*El número de observaciones para 9 de las 10 variables fue 3784; sin embargo, en la variable “Colegio Lejano” no se contaba información para 62 observaciones por lo que solo se incluyeron 3722 observaciones.

Fuente: Elaboración propia.

Las estadísticas descriptivas muestran que la nota promedio para la muestra seleccionada es de 4,27 en la base completa, de 4,48 en la de colegio cercano, y de 3,78 en la de colegio lejano, lo que indica que los estudiantes que no tienen que cambiarse de residencia para estudiar en la universidad obtienen, en promedio, calificaciones más altas. El promedio de mujeres no varía entre las bases (cerca de un 30% de los estudiantes son mujeres). Por otra parte, al igual que con la nota ordinaria, la nota de admisión promedio es más alta en los estudiantes provenientes de colegios cercanos (casi 24 puntos de diferencia entre medias) que en los de colegios lejanos.

En el resto de variables, se sigue la misma tendencia bajo la cual en Colegio Cercano los promedios son más altos que en Colegio Lejano (estas variables son: estudiantes con MATEM, de colegio privado, con laboratorio de informática); con la excepción de porcentaje de profesores titulados, que fue ligeramente superior en el caso de Colegio Lejano.

Es importante resaltar también que en el caso de los colegios cercanos, el promedio de colegios en zona de pobreza fue relativamente bajo, pues solo se encontraron 2 distritos dentro de los de mayor pobreza del país (Llano Bonito de León Cortés Castro y Uruca de San José, respectivamente). Por esta razón, se debe tener cuidado con la interpretación de esta variable en los modelos de colegio cercano, dado que la variabilidad es reducida.

Luego de la estimación de los modelos, se realizó el test de Wald para probar las diferencias entre los grupos en cada modelo, que se expondrán luego de la interpretación de los modelos.

Para la estimación de los modelos se utilizó el software MLWin, especializado en los modelos multinivel. A continuación, se presentan los resultados principales de los modelos⁶.

a. Modelo Binomial Multinivel

El cuadro 5 presenta los resultados obtenidos en los 3 modelos estimados. En el modelo de base completa, se encontró que todas las variables del nivel 1 (Mujer y Nota de Admisión) son significativas al 1%. La interpretación de los signos de los coeficientes es directa: manteniendo todo lo demás constante, el ser mujer aumenta las probabilidades de aprobar el primer curso de matemática a nivel universitario. Asimismo, a mayor nota en el examen de admisión, aumenta la probabilidad de haber aprobado este curso.

Resulta importante hacer la salvedad que el efecto de ser mujer podría obedecer a un sesgo existente por la afinidad que las estudiantes matriculadas en carreras con alta carga matemática tendrían con las matemáticas, dado que estas carreras no suelen ser demandadas de por mujeres (menos del 30% de las observaciones en la base analizada está compuesta por mujeres).

La variable MATEM resultó significativa al 1%, lo que implica que estudiantes que recibieron MATEM en el colegio tienen mayores probabilidades de aprobar el primer curso de matemática universitario. Se debe tomar en consideración que estudiantes que llevaron

⁶ Antes de realizar las estimaciones, se realizó una prueba ANOVA para comprobar que existen diferencias significativas en el rendimiento entre estudiantes de colegios cercanos y estudiantes de colegios lejanos. La prueba se muestra en el Anexo 4.

MATEM en la secundaria son alumnos interesados en carreras científicas, por lo que tendrían cierta afinidad mayor por las matemáticas.

Por otra parte, la variable Pobreza es significativa al 1%, por lo que proceder de un colegio ubicado en un distrito con altos niveles relativos de pobreza, reduce la probabilidad de aprobar el primer curso de matemática.

Estos resultados concuerdan con los efectos esperados; sin embargo, de las variables de segundo nivel, solamente Pobreza resultó ser significativa. De esta forma, en este modelo no parece haber evidencia de que provenir de un colegio privado, con laboratorio de informática y con un alto porcentaje de profesores titulados tenga efectos significativos sobre la probabilidad de aprobar el primer curso de matemática universitario.

Una variable de interés era la de Colegio Lejano, la cual resultó no ser significativa en el modelo; por lo que, al igual que las otras variables de segundo nivel, no tiene efectos en la probabilidad de aprobar.

La significancia de las variables fue consistente con la obtenida con las otras bases (colegio cercano y colegio lejano), puesto que las variables de primer nivel⁷ también resultaron significativas y con los mismos signos que en el modelo completo (Mujer dejó de serlo al 1% en el caso de colegio lejano, pero siguió siéndolo al 5%).

De igual forma, al igual que en el modelo con base completa, las variables de segundo nivel no resultaron significativas, con excepción de la variable Pobreza, la cual al 10% de significancia resulta significativa en el modelo con base de colegio lejano (Pobreza no fue incluida en el modelo de colegio cercano). De forma consistente con la literatura y con el resultado del modelo de base completa, el efecto de la variable Pobreza es negativo sobre la probabilidad de aprobar.

⁷ MATEM no fue incluida en la regresión con base de colegio Lejano porque se cuentan con muy pocas observaciones y la reducida variabilidad de éstas respecto a la variable dependiente generaría problemas de estimación y sesgos en los estimadores finales. No obstante, en el modelo multinomial sí se incluyó por mostrar mayor variabilidad.

Cuadro 5: Resultados de los Modelos Binomiales

Variable Dependiente: Aprueba

Efectos Fijos

	Base Completa			Base Colegio Cercano			Base Colegio Lejano		
	Coefficiente	Valor P	Signf.	Coefficiente	Valor P	Signf.	Coefficiente	Valor P	Signf.
Constante	-0,707	0,0000	***	-0,765	0,0000	***	-0,549	0,0019	***
Mujer	0,334	0,0001	***	0,406	0,0001	***	0,34	0,0263	**
Nota de Admisión	0,707	0,0000	***	0,854	0,0000	***	0,432	0,0000	***
MATEM	2,058	0,0000	***	1,911	0,0000	***			
Colegio Privado	0,111	0,2671		0,089	0,4730		0,05	0,7847	
Prof. Titulados	0,038	0,4188		0,039	0,5086		0,114	0,2054	
Lab. De Informática	0,117	0,2607		0,185	0,1453		0,017	0,9252	
Pobreza	-1,011	0,0000	**				-0,731	0,0754	*
Colegio Lejano	0,049	0,6343							
Distancia				0,001	0,8864				

Efectos Aleatorios

	Base Completa			Base Colegio Cercano			Base Colegio Lejano		
	Coefficiente	Valor P	Signf.	Coefficiente	Valor P	Signf.	Coefficiente	Valor P	Signf.
Constante/Constante	0,541	0,0000	***	0,482	0,0000	***	0,625	0,0001	**
Mujer/Constante	0,000			0,000			0		
Mujer/Mujer	0,000			0,000			0		
NotaAdmisión/Constante	0,032	0,5607		0,026	0,6936		0		
NotaAdmisión/Mujer	0,000			0,000			0		
NotaAdmisión/NotaAdmisión	0,159	0,0317	**	0,168	0,0709	*	0		

(***) Significativo al 1%; (**) Significativo al 5%; (*) Significativo al 10%.

Fuente: Elaboración Propia

La variable Distancia, incluida solo en el modelo con base de colegio cercano, no parece tener relación en la probabilidad de aprobar, pues no resultó ser significativa.

Los coeficientes de las variables indican el efecto en la probabilidad-logarítmica (*log-odds*) de aprobar el primer curso de matemática. Siendo así, en el modelo de base completa, en promedio, la probabilidad de aprobar es 0,334 más alta para las mujeres que para los hombres, o 0,923 veces más baja en el caso de estudiantes que provienen de zonas de pobreza.

La parte de efectos aleatorios del modelo muestra la variabilidad de los estudiantes de los colegios. El coeficiente de Constante/Constante presenta la varianza de los residuos aleatorios σ_{u0}^2 , es decir, los residuos entre el valor predicho por el modelo y el valor real a nivel de colegio, mientras que los otros coeficientes corresponden a covarianzas entre los residuos de cada coeficiente del nivel 1.

Para la parte de efectos aleatorios del modelo, se realizaron Test de Wald con el fin de determinar si los efectos de las variables de primer nivel varían dentro de los colegios, y así comprobar la segunda hipótesis de investigación. El Test de Wald prueba la hipótesis de que los coeficientes son iguales a cero, por lo que si se rechaza tal hipótesis, se concluye que dentro de los colegios del país, se tienen diferencias significativas de las variables de primer nivel en el rendimiento. Los resultados de los test se presentan en el cuadro a continuación:

Cuadro 6. Test de Wald para las variables de primer nivel. Modelo Binomial.

	Base Completa			Base Colegio Cercano			Base Colegio Lejano		
	Chi-Cuadrado	Valor P	Signf.	Chi-Cuadrado	Valor P	Signf.	Chi-Cuadrado	Valor P	Signf.
Cons-Cons	36,77	0,00	***	22,67	0,00	***	14,53	0,00	***
Cons-Mujer	0,00	~1		0,00	~1		0,00	~1	
Cons-Admis	5,11	0,08	*	3,70	0,16		0,00	~1	

(**) Significativo al 5%; (*) Significativo al 10%.

Fuente: Elaboración Propia

El componente Cons/Cons es significativa al 5% en el test de Wald (modelo completo), lo que indica que se rechaza la hipótesis nula de que “ $\sigma_{u0}^2=0$ ”, por lo que se rechaza que no existan diferencias significativas entre los colegios en el rendimiento académico. En el resto de modelos se obtuvo el mismo efecto al mismo nivel de significancia. Este resultado es de interés pues a pesar de que ciertos colegios (públicos y privados) realizan procesos de admisión, se tienen diferencias significativas en los estudiantes dentro de los colegios.

En cuanto a la variable Mujer, el test no se rechazó en ningún caso, lo que implica que no es posible afirmar que haya variaciones significativas en el rendimiento entre mujeres **dentro** de los colegios.

Por último, el efecto de la nota en el examen de admisión resultó variar dentro de los colegios en el modelo completo, pues el parámetro fue significativo al 10%. Lo que indica que a nivel nacional existen diferencias significativas en el rendimiento para estudiantes con diferentes notas de admisión dentro de un mismo colegio.

b. Modelos Multinomial Ordenado Multinivel

Los resultados de los modelos multinomiales se muestran en el cuadro 7. Las variables identificadas con “(N)” implican que la variable fue normalizada para el análisis. El primer grupo de variables viene dado por los interceptos para cada categoría, mientras que el segundo incluye las variables explicativas.

Antes de realizar el análisis e interpretación de este cuadro, es importante tener en cuenta que la categoría de referencia es “Alto” (nota de 8 o más), por lo que los coeficientes de las variables explicativas vienen en función del resto de categorías. Esta aclaración se realiza pues los coeficientes de este modelo vienen de manera inversa al modelo binomial, por lo que la interpretación es diferente. Una vez que se realiza la conversión con el *anti-logit* la interpretación es igual a la del modelo binomial, por lo que resultan comparables.

Como referencia, en el Anexo 5 se presenta un mapa con los colegios de procedencia según la cantidad de estudiantes con notas de categoría “Alta” en el país. Se puede notar como la concentración de estos colegios se da en la zona central.

Al igual que en el caso de los modelos binomiales, las variables de primer nivel resultaron significativas en todos los modelos al 1% (Mujer al 10% en el caso de colegio lejano), así como la variable Pobreza de segundo nivel.

No obstante, en el caso de las variables de segundo nivel, sí se encontraron diferencias tanto en comparación con los modelos binomiales, como entre los modelos multinomiales. La variable Colegio Privado resultó significativa en el modelo de base completa, pero no en los otros dos modelos. Por otra parte, el porcentaje de profesores titulados solo resultó significativo en el modelo con base de Colegios Lejanos, mientras que el provenir de un colegio con laboratorio de cómputo fue significativo solo en el modelo con base de Colegios Cercanos.

El hecho de que algunas variables de segundo nivel resulten significativas solo en algunos de los modelos multinomiales se puede explicar a través de la comparación entre los colegios e individuos de cada modelo, ya que una variable puede tener un efecto importante en el rendimiento dependiendo de los colegios con los que se compare. A manera de ejemplo, el efecto de provenir de un colegio privado parece generar diferencias en el rendimiento al comparar a todos los estudiantes del país, pero cuando se compara entre sólo los estudiantes de la zona central, el venir de colegio privado no es significativo; de igual forma, el porcentaje de profesores titulados parece solo tener un efecto en el rendimiento de los estudiantes de zona alejada, no así en los de zona cercana o a nivel general entre todos los estudiantes.

Al igual que en los modelos binomiales, las variables de Colegio Lejano y de Distancia no resultaron significativas.

Cuadro 7: Resultados de los modelos multinomiales

Variable Dependiente: Nota Categórica

Variable de Referencia: Alto

Efectos Fijos

	Base Completa			Base Colegio Cercano			Base Colegio Lejano		
	Coefficiente	Valor P	Signif.	Coefficiente	Valor P	Signif.	Coefficiente	Valor P	Signif.
cons.(≤Cero)	-2,055	0,000	***	-2,2620	0,0000	***	-2,443	0,0000	***
cons.(≤Bajo)	0,189	0,514		0,0450	0,7462		-0,249	0,0970	*
cons.(≤Insuficiente)	0,952	0,000	***	0,8030	0,0000	***	0,544	0,0003	***
cons.(≤Aceptable)	2,222	0,000	***	2,0890	0,0000	***	1,798	0,0000	***
Mujer	-0,257	0,000	***	-0,3080	0,0002	***	-0,230	0,0747	*
Nota Admisión (N)	-0,648	0,000	***	-0,7780	0,0000	***	-0,451	0,0000	***
MATEM	-1,602	0,000	***	-1,4280	0,0000	***	-2,287	0,0000	***
Colegio Privado	-0,147	0,070	*	-0,1690	0,1009		-0,088	0,5627	
Porc. Prof. Titulados (N)	-0,037	0,317		-0,0120	0,8026		-0,141	0,0441	**
Laboratorio de Informática	-0,136	0,106		-0,2140	0,0397	**	0,017	0,9092	
Pobreza	0,704	0,006	***				0,573	0,0629	*
Colegio Lejano	-0,123	0,138							
Distancia				-0,0040	0,5050				

Efectos Aleatorios

Nivel: Colegio

	Base Completa			Base Colegio Cercano			Base Colegio Lejano		
	Coefficiente	Valor P	Signif.	Coefficiente	Valor P	Signif.	Coefficiente	Valor P	Signif.
Constante/Constante	0,388	0,000	***	0,4020	0,0000	***	0,415	0,0002	***
Mujer/Constante	0,000			0,000			0,000		
Mujer/Mujer	0,000			0,000			0,000		
NotaAdmisión/Constante	0,049	0,1140		0,000			0,000		
NotaAdmisión/Mujer	0,000			0,000			0,000		
NotaAdmisión/NotaAdmisión	0,026	0,4702		0,000			0,000		

(***) Significativo al 1%; (**) Significativo al 5%; (*) Significativo al 10%.

Fuente: Elaboración propia

Si bien a partir de los resultados del cuadro 7 es posible interpretar relaciones entre los signos de los coeficientes, lo más recomendado en la literatura es convertir los coeficientes logit en probabilidades (anti-logit⁸). Esta conversión permite ver el cambio en las probabilidades de cada una de las cinco categorías de la variable dependiente, por lo que ya no se tiene una variable de referencia sino que todas pueden ser analizadas a la vez.

El primer paso para obtener los coeficientes en probabilidades es convertir los coeficientes de las constantes en probabilidades acumuladas. A partir de éstas, a sabiendas de que la categoría de referencia es “Alto” (Probabilidad acumulada de “Alto”=1), se pueden obtener las probabilidades condicionadas por categoría. El término “condicionada” hace referencia a que son las probabilidades obtenidas cuando solo las constantes son parte del modelo, es decir, cuando el resto de variables toman el valor de “0”. En este caso se estiman las probabilidades para un hombre (Mujer=0) con una nota de admisión promedio (Nota de Admisión Normalizada=0), sin llevar MATEM (MATEM=0), de colegio público (Privado=0), con un porcentaje de profesores titulados promedio (Porc. Prof Titulados Normalizada=0), sin Laboratorio de Informática (Lab Informática=0), en zona de no pobreza (Pobreza=0) y proveniente de un colegio cercano (Colegio Lejano=0). Los resultados de la estimación se presentan en el cuadro 8 en términos de probabilidades condicionadas.

Cuadro 8: Probabilidades Condicionadas de los modelos multinomiales

Categoría	Base Completa		Base Colegio Cercano		Base Colegio Lejano	
	Probabilidad Acumulada	Probabilidad	Probabilidad Acumulada	Probabilidad	Probabilidad Acumulada	Probabilidad
Cero	0,09	0,09	0,09	0,09	0,08	0,08
Bajo	0,49	0,39	0,51	0,42	0,44	0,36
Insuficiente	0,67	0,18	0,69	0,18	0,63	0,19
Aceptable	0,88	0,21	0,89	0,20	0,86	0,23
Alto	1,00	0,12	1,00	0,11	1,00	0,14

Fuente: Elaboración propia

Si se quisiera conocer el efecto de ser mujer (Mujer=1), las probabilidades incorporarían ahora el coeficiente de la variable Mujer en el antilogit, por lo que las nuevas probabilidades

⁸ El anti-logit viene dado por $\pi_y = \frac{e^x}{1+e^x}$

vendrían dadas por el Cuadro 9, el cual incluye el cambio porcentual respecto al primer grupo de probabilidades (Cuadro 8):

Cuadro 9: Probabilidades Condicionadas para Mujer de los modelos multinomiales

Categoría	Base Completa		Base Colegio Cercano		Base Colegio Lejano	
	Probabilidad	Cambio Porcentual	Probabilidad	Cambio Porcentual	Probabilidad	Cambio Porcentual
Cero	0,07	-0,02	0,07	-0,02	0,06	-0,02
Bajo	0,35	-0,04	0,36	-0,06	0,32	-0,04
Insuficiente	0,19	0,01	0,19	0,01	0,20	0,01
Aceptable	0,24	0,03	0,23	0,03	0,25	0,02
Alto	0,15	0,03	0,14	0,03	0,17	0,03

Fuente: Elaboración propia

De esta forma, manteniendo todo lo demás constante, las probabilidades de tener una nota alta o aceptable aumentan para las mujeres en todos los casos, mientras que las probabilidades de tener una nota de cero o baja, se reducen, respecto al primer grupo.

El Cuadro 10 muestra los cambios porcentuales respecto al primer grupo de probabilidades (Cuadro 8) para las variables que se encontraron significativas en el modelo. Como se muestra, el aumento de una desviación estándar en la nota normalizada de admisión obtenida aumenta las probabilidades de tener una nota alta o aceptable, y disminuye las de tener un cero de nota final. Además, el efecto parece ser mayor en el caso de los colegios cercanos (la probabilidad de tener una nota alta aumenta en 10 puntos porcentuales).

Cuadro 10: Cambios en las probabilidades condicionadas para las variables significativas de los modelos.

	Nota de Admisión			MATEM		
	Completo	Ccercano	Clejano	Completo	Ccercano	Clejano
Cero	-0,04	-0,04	-0,03	-0,07	-0,07	-0,07
Bajo	-0,11	-0,14	-0,08	-0,25	-0,24	-0,3
Insuficiente	0,00	0,00	0,00	-0,05	-0,03	-0,11
Aceptable	0,07	0,08	0,04	0,09	0,11	0,00
Alto	0,09	0,1	0,07	0,29	0,23	0,48

	Colegio Privado			Porc. Prof. Titulados		
	Completo	Ccercano	Clejano	Completo	Ccercano	Clejano
Cero	-0,01					-0,01
Bajo	-0,02					-0,03
Insuficiente	0,01					0,01
Aceptable	0,02					0,01
Alto	0,02					0,02

	Laboratorio de Informática			Pobreza		
	Completo	Ccercano	Clejano	Completo	Ccercano	Clejano
Cero		-0,01		0,08		0,05
Bajo		-0,04		0,10		0,09
Insuficiente		0,01		-0,03		-0,02
Aceptable		0,02		-0,08		-0,07
Alto		0,02		-0,06		-0,05

Fuente: Elaboración propia

Por otra parte, la variable en la que se presentan los mayores cambios en las probabilidades es la de haber llevado MATEM en el colegio, pues estudiantes de colegios que ofrecían MATEM tienen menos probabilidades de reprobado (es decir, de tener notas de cero, bajas o insuficientes) y más del doble de probabilidades de aprobar (notas aceptables y altas). Este efecto es especialmente mayor en los colegios lejanos.

Se aprecia que el efecto más fuerte se encontró en la variable MATEM, en especial en los colegios de zonas alejadas. Encontrar este tipo de resultados es de relevancia para el diseño de políticas que ayuden a mejorar el rendimiento académico de los estudiantes. No obstante, se debe tener en consideración que puede haber un sesgo de selección en esta variable, puesto que los estudiantes que tienen acceso y matriculan MATEM, tienen afinidad por las matemáticas, por lo que MATEM resultaría un indicador de las destrezas en matemáticas, y no solo la causa de tales.

El efecto de las variables de segundo nivel es el esperado, pues Colegio Privado aumenta (ligeramente) las probabilidades de obtener una nota alta o aceptable, así como el porcentaje de profesores titulados y el contar con laboratorio de cómputo.

Por otra parte, Pobreza tiene un efecto contrario en la probabilidad de aprobar. El provenir de colegios en zona de pobreza reduce las probabilidades de obtener una nota alta, aceptable e incluso insuficiente, mientras que eleva la de tener una nota de cero o baja.

Al igual que con los modelos binomiales, también se realizaron las pruebas de Wald para los efectos aleatorios en los modelos multinomiales. Los resultados para los tres modelos se muestran en el Cuadro 11. En este caso, para todos los modelos, el parámetro Cons/Cons tuvo un valor significativo al 1% en el test de Wald, con lo que se rechaza la hipótesis de igualdad de rendimiento dentro de los estudiantes de un mismo colegio.

Cuadro 11. Test de Wald para las variables de primer nivel. Modelo Multinomial.

	Base Completa			Base Colegio Cercano			Base Colegio Lejano		
	Chi-Cuadrado	Valor P	Signf.	Chi-Cuadrado	Valor P	Signf.	Chi-Cuadrado	Valor P	Signf.
Cons-Cons	41,832	0,000	***	33,454	0,000	***	13,679	0,000	***
Cons-Mujer	0,000	~1		0,000	~1		0,000	~1	
Cons-Admis	2,791	0,248		0,000	~1		0,000	~1	

(***) Significativo al 1%; (**) Significativo al 5%; (*) Significativo al 10%.

Fuente: Elaboración propia

Tanto en el caso de la variable Mujer como en la de Nota de Admisión, no se rechazó la hipótesis nula de igualdad de rendimiento entre estudiantes, por lo que no se encontró evidencia de diferencias significativas asociadas al efecto de ser Mujer y a la Nota de Admisión en el rendimiento entre estudiantes de un mismo colegio.

La matriz del cuadro 12 presenta un resumen con las variables que resultaron significativas en cada modelo y el efecto que tienen sobre el rendimiento en el primer curso de matemática, aproximado por la probabilidad de aprobar. Para homologar el análisis, en el caso de los

modelos multinomiales, se consideró que existe un “efecto positivo” en el rendimiento si la variable aumenta la probabilidad de obtener una nota Aceptable o Alta.

Cuadro 12. Resumen de variables significativas y su efecto en el rendimiento.

	Base Completa		Base Colegio Cercano		Base Colegio Lejano	
	Binomial	Multinomial	Binomial	Multinomial	Binomial	Multinomial
Mujer	+	+	+	+	+	+
Nota Admisión	+	+	+	+	+	+
MATEM	+	+	+	+	+	+
Colegio Privado		+				
Porc. Prof. Titulados						+
Laboratorio de Informática				+		
Pobreza	-	-			-	-
Colegio Lejano Distancia						

Fuente: Elaboración propia

El cuadro 13 muestra los resultados de las pruebas de Wald para cada modelo estimado.

Cuadro 13: Resumen significancia de la prueba de Wald

	Base Completa		Base Colegio Cercano		Base Colegio Lejano	
	Binomial	Multinomial	Binomial	Multinomial	Binomial	Multinomial
Dentro de los Colegios Mujeres	***	***	***	***	***	***
Nota Admisión	*					

(***) Significativo al 1%; (**) Significativo al 5%; (*) Significativo al 10%.

Fuente: Elaboración Propia

Conclusiones

Luego del análisis de los modelos econométricos realizados, se encuentra que los resultados rechazan la primera hipótesis de investigación, puesto que estudiantes provenientes de colegios con características distintas tienen rendimientos significativamente distintos en el primer curso de matemática universitaria.

Estudiantes de colegios públicos muestran diferencias en el rendimiento académico en comparación con los de colegios privados, hecho que se podría explicar por los procesos de admisión que algunos colegios privados aplican a sus estudiantes. De igual forma, estudiantes provenientes de colegios con laboratorio de informática o con diferentes porcentajes de profesores titulados presentan rendimientos significativamente distintos.

Los resultados de los tests de Wald rechazaron en los 6 modelos la hipótesis nula, con lo que se rechaza la segunda hipótesis de investigación, estudiantes provenientes de un mismo colegio tienen el mismo rendimiento académico. Esto además ratifica el uso de modelos multinivel en este tipo de análisis, pues existen diferencias significativas en el rendimiento dentro de estudiantes de un mismo colegio, por lo que utilizar modelos de un solo nivel implicaría sesgos al pensar que todos los estudiantes del país son observaciones aleatorias de la población.

Las mujeres tienen mayores probabilidades de presentar un rendimiento alto en el primer curso de matemática, no obstante, esta diferencia no se presenta a nivel colegial (entre estudiantes de un mismo colegio), sino a nivel agregado entre todos los estudiantes de la base de datos. Lo mismo ocurre con el capital cultural del estudiante, aproximado por su nota de admisión en la universidad, se presentan diferencias en el rendimiento a nivel general, pero no dentro de los estudiantes de un mismo colegio, lo que se interpretaría como uniformidad en tales variables.

Los efectos de las variables geográficas específicas no presentaron coeficientes significativos en ninguno de los modelos, ni el provenir de zona lejana, ni la distancia de desplazamiento en el caso de los estudiantes de colegio cercano.

Por otra parte, al realizar las estimaciones separando los modelos según la localización del estudiante antes de entrar a la universidad, se encontró que los efectos de las variables de segundo nivel varían según la ubicación espacial. En específico, el porcentaje de profesores titulados tiene un efecto positivo y significativo en los estudiantes de zonas lejanas, mientras que el tener laboratorio de cómputo fue significativo solo para los de zonas cercanas.

Así entonces, la tercera hipótesis de investigación: las variables del contexto colegial (características del colegio de procedencia) mantienen su significancia en el rendimiento académico sin importar la ubicación geográfica del colegio, también se rechaza, puesto que se encontró que la significancia de las variables del colegio de procedencia varía según la ubicación geográfica del colegio (si está en zona cercana o en zona lejana).

Se debe considerar que en el estudio solo se incluyeron estudiantes admitidos en carreras de alta carga matemática que matricularon el curso, por lo que podría haber un “sesgo de selección” en cuanto a que los estudiantes incluidos ya han pasado por varios filtros y podrían poseer características intrínsecas que los diferencian positivamente de otros estudiantes.

Otra limitación del estudio fue la escasa inclusión de variables de índice social. La mayor dificultad al respecto es encontrar variables sociales específicas de los colegios o de los estudiantes sin llevar a cabo un estudio de campo. Una alternativa a esta problemática sería obtener acceso a los datos socioeconómicos de los estudiantes que la Oficina de Becas de la UCR administra (al menos para los estudiantes que solicitaron beca).

Aunado a esto, un análisis de interés sería el incluir una nueva variable por carrera o curso para explorar si hay diferencias según la carrera o curso matriculado por el estudiante, puesto que en esta investigación los estudiantes de todas las carreras con alta carga matemática se trataron por igual para los tres cursos seleccionados.

Si bien en esta investigación se tenía como objetivo ver el efecto socio-geográfico del colegio de procedencia en el primer curso de matemática, un punto de interés sería estudiar el rendimiento total en todos los cursos de matemática en los que el estudiante se matricule, con lo que se podría incluir si el estudiante desistió o si logró mejorar su rendimiento, y si las diferencias que podrían haber habido en un principio entre estudiantes, se logran corregir con el tiempo que el estudiante acumula en la universidad.

A pesar de las limitaciones o mejoras que se podrían realizar a la presente investigación, se tiene un avance exploratorio importante sobre los determinantes del rendimiento académico en matemática universitaria de la UCR, de interés por las implicaciones en política que este tema

conlleva en dos dimensiones: a nivel universitario y a nivel nacional. La investigación en rendimiento académico brinda insumos que pueden usarse para la toma de decisiones en el campo, por lo que se debe fomentar el uso de técnicas innovadoras para el país y contribuir a la literatura empírica del tema en Costa Rica.

Por el lado universitario, el constante fracaso en el primer curso de matemática no es un tema nuevo para las escuelas de matemática de las universidades públicas del país, pero continúa siendo un obstáculo para el buen desempeño de los estudiantes en las distintas carreras con alta carga matemática. A nivel universitario, la inclusión del DIMA como examen obligatorio y los cursos de matemática elemental y pre-cálculo se espera que mejoren los resultados en las asignaturas de la escuela de matemática, pero realizar un análisis de tales medidas requiere más tiempo desde la fecha de implementación.

A nivel nacional, resulta de interés como el país invierte más en educación per cápita que los países de la OECD, más los resultados en pruebas como PISA son más bajos en comparación. Estas diferencias deben ser focos de atención si se considera que el país busca tener una población educada y calificada para puestos de alto nivel de conocimiento. A nivel colegial, el MEP ha realizado esfuerzos por mejorar el rendimiento en matemática, no obstante, un análisis del efecto de tales medidas solo se podría realizar a mediano o largo plazo, puesto que son de reciente implementación.

Referencias

- Agüero, M., & Castillo, D. (19 de 03 de 2015). *nacion.com*. Recuperado el 16 de 07 de 2015, de http://www.nacion.com/data/Alumnos-carreras-hallan-empleo-aulas_0_1476252361.html
- Alpizar, M., & Alfaro, A. L. (2013). EL PROYECTO MATEMÁTICA PARA LA ENSEÑANZA MEDIA (MATEMÁTICA): PERCEPCIÓN DE LOS ESTUDIANTES SOBRE LOS CURSOS RECIBIDOS Y LAS CARRERAS EN EDUCACIÓN SUPERIOR QUE ELIGIERON. *UNICIENCIA Vol. 27, N° 1*, 34-58.
- Barragán, M. (2008). *El rendimiento académico de los estudiantes de educación superior: el caso de la Universidad Autónoma de Guerrero, México*. San José: Universidad de Costa Rica. Sistema de Estudios de Posgrado.
- Becker, G. (1964). *Human Capital*. New York: Columbia University Press for the National Bureau of Economic Research.
- Card, D., & Krueger, A. (1992). Does School Quality Matter? Returns to Education and the Characteristics of Public School in the United States. *Journal of Political Economy*, *100 (1)*, 1-40.
- Cheers, B. (1990). Rural disadvantage in Australia. *Australian Social Work*, *Vol. 43, No. 1*, 5-13.
- Contreras, D., Gallegos, S., & Meneses, F. (2009). University performance determinants: does relative ability matter? *Calidad en la Educación N° 30*, 18-48.
- Cordero, C. (21 de 01 de 2016). *El Financiero*. Recuperado el 16 de 08 de 2016, de http://www.elfinancierocr.com/tecnologia/Intel-Timothy_Scott-empleo-big_data-cloud_0_888511144.html
- Dickey, S., & Houston, R. (2010). Estimation of the Education Production Function for Principles for Macroeconomics. *Perspectives on Economic Education Research*, 72-89.
- Falch, T., Lujala, P., & Strom, B. (2011). *Geographical constraints and educational attainment*. Norway: Norwegian University of Science and Technology.
- Fernández, A., & Del Valle, R. (2013). Desigualdad educativa en Costa Rica: la brecha entre estudiantes de colegios públicos y privados. Análisis con los resultados de la evaluación internacional PISA. *CEPAL 111*, 37-57.
- Fernández, E. (15 de 12 de 2013). *Pruebas de PISA muestran que Costa Rica tiene una fórmula matemática por resolver*. Recuperado el 14 de 08 de 2016, de El Financiero: http://www.elfinancierocr.com/economia-y-politica/pruebas_PISA-educacion_de_Costa_Rica-leonardo_garnier_0_427157313.html
- Friedman, M. (1961). *Capitalismo y Libertad*. Chicago: University Of Chicago.
- Garbanzo, G. (2014). Factores asociados al rendimiento académico tomando en cuenta el nivel socioeconómico: Estudio de regresión múltiple en estudiantes universitarios. *Revista Electrónica Educare Vol 18 (1)*, 119-154.

- García, F., & Doménech, F. (1997). Motivación, Aprendizaje y Rendimiento Escolar. *Revista Electrónica de Motivación y Emoción*, N° 1 (0), Recuperado de <http://reme.uji.es/articulos/pa0001/texto.html>.
- Goldstein, H. (2003). *Multilevel statistical models*. London: Arnold.
- Instituto Nacional de Estadística y Censos. (2013). *Costa Rica: Mapas de Pobreza 2011*. San José: INEC.
- Johnson, I. (2008). Enrollment, Persistence and Graduation of In-State Students at a Public Research University: Does High School Matter? *Research in Higher Education*, 776-793.
- López, R. (2009). College Proximity: Mapping Access to Opportunity. *Sociology of Education*, Vol. 82, No. 2, 126-146.
- Lucas, R. (1988). *On the Mechanics of Economic Development*. Chicago: University of Chicago.
- Martínez, V. (2002). Condicionantes del rendimiento escolar. *Educadores Revista de Renovación Pedagógica* N° 47 (204), 285-295.
- Mora, L. (1998). *La educación como proceso integral correlacionado: análisis de la transición entre el ciclo diversificado y la educación universitaria*. San José: Universidad de Costa Rica. Sistema de Estudios de Posgrado. Escuela de Administración Educativa.
- Moreira Mora, T. (2013). Factores de contexto, entrada y proceso asociados al rendimiento en Matemática: un estudio multinivel. *Actualidades en Psicología*, Vol N° 27 (114), 19-38.
- Öckert, B. (2012). On the margin of success? Effects of expanding higher education for marginal students. *Nordic Economic Policy Review*, No. 1, 111-159.
- Ortega, J. A. (2001). Revisión de conceptos demográficos. *Contribuciones a la Economía*, Recuperado de <http://www.eumed.net/coursecon/colaboraciones/Ortega-demograf.htm>.
- Oviedo, Y. (2012). *Factores asociados al rendimiento académico en Matemática en el III ciclo de la Educación General Básica; Un estudio multinivel*. San José: Estado de la Educación.
- Programa Estado de la Nación. (2013). *Quinto Informe Estado de la Educación*. San José: Programa Estado de la Nación.
- Ramírez, B. (1993). *Éxito alcanzado en la educación superior por estudiantes que egresaron de colegios privados, en relación con los estudiantes que egresaron de colegios públicos y subvencionados*. San José: Universidad de Costa Rica. Comisión del Programa de Estudios de Posgrado en Educación.
- Rasbash, J., Steele, F., Browne, W., & Goldstein, H. (2015). *A User's Guide to MLWin*. Bristol: Centre for Multilevel Modelling, University of Bristol.
- Rephann, T. J. (1996). The Importance of Geographical Attributes in the Decision to Attend College. *Socio-Economic Planning Sciences*, Vol. 36, No. 4, 291-307.
- Schultz, T. (1961). Investment in Human Capital. *The American Economic Review*, Vol. 41, N° 1, 1-17.
- Trusty, J., & Niles, S. (2003). High-School Math Courses and Completion of the Bachelor's Degree. *Professional School Counseling*, 99-107.
- Universidad de Costa Rica. (2001). *Reglamento de Régimen Académico Estudiantil*. San José: Universidad de Costa Rica.
- Valverde, A. (2005). *Factores asociados al rendimiento académico en la UCR: un modelo estructural de variables latentes*. San José: Universidad de Costa Rica. Sistema de Estudios de Posgrado.

- Vargas, M. (2010). *Factores que determinan el rendimiento académico en matemáticas en la Universidad Nacional de Ingeniería, Nicaragua: un estudio multinivel y de ecuaciones lineales estructurales*. San José: Universidad de Costa Rica. Sistema de Estudios de Posgrado.
- Vásquez Olivera, M. (2015). La calidad de la educación: Reformas educativas y control social en América Latina. *Temas y Problemas de Nuestra América*, 93-124.
- Vicerrectoría de Docencia. (2009). *RESOLUCIÓN VD-R-8428-2009*. San Pedro de Montes de Oca: Universidad de Costa Rica.

Anexos

Anexo 1: Distribución de Frecuencias de las variables dependientes de los modelos

	Categoría	Completo		Ccercano		Clejano	
		Frecuencia Absoluta	Frecuencia Relativa	Frecuencia Absoluta	Frecuencia Relativa	Frecuencia Absoluta	Frecuencia Relativa
Modelo Binomial	Reprueba	2113	56,8%	1427	53,9%	686	63,8%
	Aprueba	1609	43,2%	1219	46,1%	390	36,2%
	Total	3722		2646		1076	
Modelo Multinomial	Cero	311	8,4%	202	7,6%	109	10,1%
	Bajo	1202	32,3%	816	30,8%	386	35,9%
	Insuficiente	600	16,1%	409	15,5%	191	17,8%
	Aceptable	841	22,6%	624	23,6%	217	20,2%
	Alto	768	20,6%	595	22,5%	173	16,1%
	Total	3722		2646		1076	

Fuente: Elaboración propia

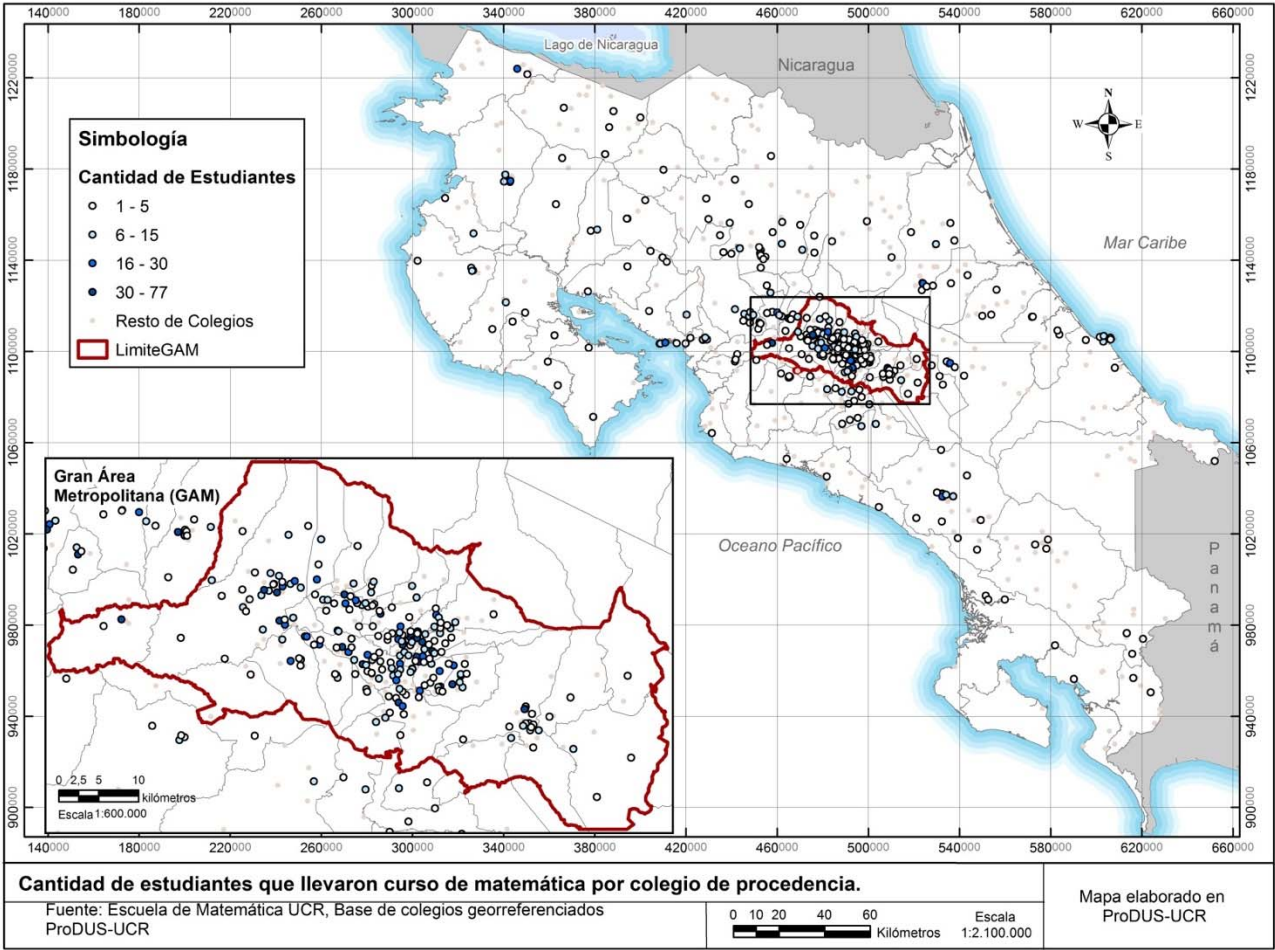
Anexo 2: Matriz de Correlaciones de las variables del Modelo

	NotaOrdinaria	Aprueba	Mujer	NotaAdmision	MATEM	DIMA	CPrivado	CTitulados	CInformatica	Cdistancia	Clejano	Pobreza
NotaOrdinaria	1											
Aprueba	0,94	1										
Mujer	0,05	0,05	1									
NotaAdmision	0,35	0,29	-0,02	1								
MATEM	0,18	0,16	0,00	0,15	1							
DIMA	0,18	0,17	-0,01	0,03	0,21	1						
CPrivado	0,12	0,10	0,02	0,22	0,07	0,07	1					
CTitulados	0,01	0,02	0,02	0,00	0,00	-0,02	-0,15	1				
CInformatica	-0,03	-0,02	-0,01	-0,13	-0,12	0,14	-0,19	0,19	1			
Cdistancia	-0,08	-0,08	0,00	-0,17	-0,06	-0,08	-0,29	0,03	-0,03	1		
Clejano	-0,10	-0,09	0,00	-0,18	-0,08	-0,09	-0,30	0,03	-0,04	0,78	1	
Pobreza	-0,09	-0,08	-0,02	-0,05	-0,02	-0,05	-0,14	-0,04	-0,01	0,29	0,24	1

Énfasis en las correlaciones mayores a 0,3 en nivel absoluto.

Fuente: Elaboración propia

Anexo 3: Mapa por colegio de procedencia, según cantidad de estudiantes en la base de datos



Anexo 4: Análisis de Variancia Colegio Lejano en Nota Ordinaria

	Grados de Libertad	Suma de Cuadrados	Cuadrado Medio	F
Entre Grupos	1	20.116	20.116	11.93
Dentro de los grupos	3720	6274.4	16.867	
Total	3721	6294.5	16.916	

Fuente: Elaboración propia

Anexo 5: Mapa de colegios de procedencia por cantidad de estudiantes con nota categoría "Alta"

