

DISEÑO DE UN MODELO ESTOCÁSTICO USANDO CADENAS DE MARKOV
PARA PRONOSTICAR LA DESERCIÓN ACADÉMICA DE ESTUDIANTES DE
INGENIERÍA. CASO: ESCUELA COLOMBIANA DE INGENIERÍA JULIO GARAVITO

CLAUDIA YADIRA RODRÍGUEZ RÍOS

PONTIFICIA UNIVERSIDAD JAVERIANA
MAESTRÍA EN INGENIERÍA INDUSTRIAL
BOGOTÁ, D.C.
2012

DISEÑO DE UN MODELO ESTOCÁSTICO USANDO CADENAS DE MARKOV
PARA PRONOSTICAR LA DESERCIÓN ACADÉMICA DE ESTUDIANTES DE
INGENIERÍA. CASO: ESCUELA COLOMBIANA DE INGENIERÍA JULIO GARAVITO

CLAUDIA YADIRA RODRÍGUEZ RIOS

Trabajo de grado para optar al título
de magíster en Ingeniería Industrial

DIRECTORA
ZULIMA ORTIZ BAYONA

PONTIFICIA UNIVERSIDAD JAVERIANA
MAESTRÍA EN INGENIERÍA INDUSTRIAL
BOGOTÁ, D.C.
2011

AGRADECIMIENTOS

A Dios, pues para Él nada es imposible...

A la Escuela Colombiana de Ingeniería Julio Garavito, a través de sus directivos y de los responsables de las áreas administrativas, por el apoyo dado para este proyecto y por la colaboración prestada en todos los niveles tanto para la consecución de los datos y de la información, así como por el tiempo que se dispuso para mis estudios.

A la Dra. Zulima Ortiz, por todos sus conocimientos, por transmitirme ese amor hacia las matemáticas, por su inmenso apoyo en esta tesis, por el tiempo que me dedicó y por toda su disposición para que este trabajo tuviera los mejores resultados.

Al director de la maestría, el ing. Carlos Fúquene, una persona transparente, quien me apoyó todo el tiempo durante mis estudios.

A mis profesores de la maestría, personas íntegras que nos supieron transmitir todos sus conocimientos; hoy puedo decir que aprendí muchísimo de ustedes, que el programa es excelente y que me siento muy orgullosa de ser egresada de esta maestría.

A mi amado esposo, por haber sido mi sostén en todo este proceso de la maestría; él ha sido mi profe de probabilidad y compañero incansable, siempre animándome y dándome su amor en los momentos difíciles.

A mis hijos, David y Alejandro, por su inmensa paciencia y por el tiempo que les he quitado en la realización de este proyecto.

A mis padres, que me han apoyado en todo momento, por sus oraciones y su amor.

A Nohorita, amiga, niñera de mis hijos y en los últimos días mía también, quien ha cuidado mis hijos como si fueran suyos, e incluso me ha remplazado en muchos momentos en mi papel de madre.

DEDICADO

A Dios

A la memoria de mi querida hermana, Charito

A mi familia

CONTENIDO

LISTA DE TABLAS	ii
LISTA DE FIGURAS	iii
LISTA DE ANEXOS	iv
GLOSARIO	v
RESUMEN	1
TÍTULO.....	2
1. INTRODUCCIÓN	2
2. PLANTEAMIENTO DEL PROBLEMA	5
2.1. Antecedentes del problema.....	5
2.2. Justificación de la investigación	8
3. FORMULACIÓN DE LA PREGUNTA DE INVESTIGACIÓN	9
4. OBJETIVOS	9
5. ESTADO DEL ARTE	10
5.1. Deserción.....	10
5.2. Cadenas de Markov.....	15
6. METODOLOGÍA DE LA INVESTIGACIÓN	18
6.1. Población y muestra	18
6.2. Diseño del estudio	19
6.3. Variable Y_n	21
6.4. Variables independientes.....	23
6.5. Procedimientos	25
7. RESULTADOS	26
7.1. Reducción y limpieza de los datos	26
7.2. Recodificación de los datos	27
7.3. Integración de los datos.....	27
7.4. Estadística descriptiva de los datos	27
7.5. Análisis de valores perdidos.....	32
7.6. Prueba de independencia Chi-cuadrado	35
7.7. Programa para calcular la variable dependiente	36
7.8. Regresión logística multinomial.....	37
7.9. Matrices de transición de Markov	38
7.10. Análisis de los resultados	48
8. CONCLUSIONES	48
9. RECOMENDACIONES.....	50
10. REFERENCIAS BIBLIOGRÁFICAS	51

LISTA DE TABLAS

Tabla 1 Frecuencias iniciales de la variable dependiente.	21
Tabla 2 Matriz de probabilidades iniciales de la variable dependiente.	21
Tabla 3 Variables independientes del modelo. Elaboración propia.	24
Tabla 4 Variables administrativas del modelo. Elaboración propia.	25
Tabla 5 Descripción de los estudiantes por cohorte. Elaboración propia.	27
Tabla 6 Frecuencias y porcentajes de participación de las variables categóricas del estudio.	28
Tabla 7 Estadísticos descriptivos de las variables continuas del estudio.	28
Tabla 8 Correlaciones de las variables continuas del estudio.	29
Tabla 9 Resultados factor de significación del modelo.	29
Tabla 10 Varianza explicada por los factores.	30
Tabla 11 Factores iniciales que explican el modelo.	31
Tabla 12 Prueba de Kolmogorov-Smirnov para la normalidad de los datos.	31
Tabla 13 Prueba de Levene. Homogeneidad de los datos.	32
Tabla 14 Datos perdidos de la base de datos.	32
Tabla 15 Especificaciones de la imputación.	34
Tabla 16 Restricciones de las imputaciones.	34
Tabla 17 Resultados de la imputación.	34
Tabla 19 Matriz de probabilidades de transición, perfil 1.	39
Tabla 20 Probabilidades de transición, perfil 1, en Excel.	39
Tabla 21 Matriz M para el primer perfil analizado.	39
Tabla 22 Matriz de probabilidades de transición, perfil 2.	40
Tabla 23 Prueba Tamhane para el nivel de educación del responsable del estudiante.	42
Tabla 24 Frecuencias de la variable estrato socioeconómico.	43
Tabla 25 Prueba Tamhane para el estrato socioeconómico.	43
Tabla 26 Prueba Tamhane para la edad al entrar.	45
Tabla 27 Prueba Tamhane para con quién vive el estudiante.	46
Tabla 28 Prueba Tamhane para el tipo de vivienda del estudiante.	47
Tabla 29 Betas de la regresión logística multinomial. Resumen.	54
Tabla 30 Matrices de transición perfil 3.	56

LISTA DE FIGURAS

Figura 1 Estado del arte de los aspectos de la deserción estudiantil (Castaño et al., 2004).....	6
Figura 2 Modelo de Fishbein y Ajzen (Ajzen & Fishbein, 1977).....	12
Figura 3 Modelo de Ethington (Ethington, 1990).....	12
Figura 4 Modelo sociológico de deserción Spady (Spady, 1970).....	13
Figura 5 Modelo de integración de Tinto (Tinto, 1993).	14
Figura 6 Ejemplo de una notación gráfica de una matriz de transición.	16
Figura 7 Cadena de Markov de los estudiantes semestre a semestre. Elaboración propia.....	19
Figura 8 Matriz de transición de semestres. Elaboración propia.	20
Figura 9 Resumen global de datos perdidos.	33
Figura 10 Edad y valor de matrícula imputados.	35
Figura 11 Icfes de matemáticas, física y lenguaje imputados.....	35

LISTA DE ANEXOS

Anexo 1	Resultados regresión logística por semestres.	54
Anexo 2	Operaciones con las matrices de transición de los diferentes perfiles.	56
Anexo 3	Programa en Visual Basic Studio 2008.	58
Anexo 4	Descripción de los archivos iniciales.	65
Anexo 5	Reducción y limpieza de datos.	65
Anexo 6	Recodificación de variables.	65
Anexo 7	Deducción detallada del modelo de la regresión logística multinomial.	70

GLOSARIO

DANE: Departamento Administrativo Nacional de Estadística. Es la entidad responsable de la planeación, levantamiento, procesamiento, análisis y difusión de las estadísticas oficiales de Colombia.

Icfes: Instituto Colombiano para la Evaluación de la Educación, entidad especializada en ofrecer servicios de evaluación de la educación en todos sus niveles, y en particular apoyar al Ministerio de Educación Nacional en la realización de los exámenes de Estado y en hacer investigaciones sobre los factores que inciden en la calidad educativa.

MEN: Ministerio de Educación Nacional República de Colombia. Tiene como misión lograr una educación de calidad, que forme mejores seres humanos, ciudadanos con valores éticos, competentes, respetuosos de lo público, que ejercen los derechos humanos, cumplen con sus deberes y conviven en paz. Una educación que genere oportunidades legítimas de progreso y prosperidad para ellos y para el país. Lograr una educación competitiva, pertinente, que contribuya a cerrar brechas de inequidad y en la que participa toda la sociedad.

Sinies: Sistema Nacional de Información de la Educación Superior, creado para responder a las necesidades de información de la educación superior en Colombia. Es una fuente de información, en relación con las instituciones y los programas académicos aprobados por el Ministerio de Educación Nacional, que consolida y suministra datos, estadísticas e indicadores.

Spadies: Sistema de Información para la Prevención de la Deserción de la Educación Superior, basado en un modelo estadístico de supervivencia, que consolida y ordena información sobre los estudiantes de diferentes universidades e instituciones de educación del país.

Tasa de cobertura: Índice que señala el número de estudiantes que se matriculan en el sistema de educación nacional.

RESUMEN

Se realizó un modelo matemático estocástico usando cadenas de Markov para pronosticar la deserción académica de estudiantes de ingeniería, aplicado en un estudio de caso en la Escuela Colombiana de Ingeniería Julio Garavito; se utilizaron las cohortes del 2003 al 2006. Debido a la distribución de los datos, a que varias variables independientes son categóricas y a que la variable dependiente es una variable categórica ordenada, se empleó el modelo de regresión logística multinomial para obtener las probabilidades de transición, las que a su vez se usaron para modelar las cadenas de Markov absorbentes. Estas se utilizaron como una herramienta predictiva para determinar el tiempo de permanencia en la universidad, la probabilidad de graduarse, la probabilidad de deserción, e identificar las variables críticas que causan la deserción.

Como una de las conclusiones, se encontró que las variables independientes varían al predecir la deserción, dependiendo del semestre académico, lo que indica que no se deben generalizar las variables que afectan la deserción en todos los niveles. Variables como edad, con quién vive y género del estudiante (hombre o mujer) van cambiando su efecto a través de los semestres.

Se ha aplicado y probado un modelo matemático. La herramienta propuesta establece un valor de probabilidad de deserción y un tiempo de permanencia, los cuales pueden ayudar en la toma de decisiones de los administrativos y así prever los efectos de algunas variables que sean de control de la institución, o variables que se puedan determinar en los procesos de admisión.

Este es un estudio de caso cuantitativo y el diseño de la investigación es no experimental, longitudinal de evolución de grupo o cohorte, porque se toman los datos de cohortes completas y se analiza su comportamiento a lo largo del tiempo, aplicado a los programas de Ingeniería de la Escuela Colombiana de Ingeniería. Para utilizarlo en otras universidades se debe adaptar a la forma de calcular los semestres de acuerdo con cada institución, de tal manera que se asegure cumplir con las propiedades de las cadenas absorbentes de Markov.

TÍTULO

Diseño de un modelo estocástico usando cadenas de Markov para pronosticar la deserción académica de estudiantes de ingeniería. Caso: Escuela Colombiana de Ingeniería Julio Garavito

1. INTRODUCCIÓN

En la última década se ha presentado un incremento en la población colombiana que accede al sistema de educación superior. El número de personas matriculadas en un primer curso en el año 2000 fue de 213.529 y pasó a 488.395 estudiantes matriculados en el año 2011 (Sinies, 2012)¹, lo que representa un incremento del 128,7%. Igualmente, las tasas de cobertura² han crecido en diez puntos, pasando del 25,6% en 2003 al 35,5% en 2009 (MEN, 2010), situación que ha sido propiciada y apoyada por el gobierno nacional, que ve la educación como una herramienta para mejorar la productividad, la innovación y, por ende, la competitividad del país. Sin embargo, este incremento se ha visto afectado por la tasa de deserción, que fue de un 50% para el año 2011 (Spadies, 2012)³; es decir, por cada dos personas que ingresan, una se retira en el transcurso de sus estudios y no logra graduarse.

El problema de la deserción universitaria tiene costos sociales y económicos que afectan a las familias, a las instituciones y al gobierno. También tiene efectos negativos en la cobertura, puesto que los altos índices de deserción y un mayor tiempo de duración en obtener el grado pueden incidir en la disponibilidad de cupos para nuevos estudiantes, disminuyendo la oportunidad de acceso a estos al sistema educativo.

¹ Sinies: Sistema Nacional de Información de la Educación Superior.

² Tasas de cobertura. Índice que tiene el gobierno nacional para medir el número de personas que se matriculan en el sistema de educación superior.

³ Spadies: Sistema de información que recoge información de las universidades e instituciones técnicas para dar un valor sobre la deserción en Colombia.

Como meta para el 2019, el gobierno pretende llegar a una tasa del 25% en la deserción y lograr un 50% en el índice de cobertura⁴ (MEN, 2010); como una estrategia ante estas metas, determinó diseñar e implementar una metodología de seguimiento a la deserción, el cual denominó Sistema de Prevención de la Deserción en Educación Superior (Spadies), basado en el modelo estadístico de la supervivencia, que busca establecer una visión nacional sobre el tema, facilitar la documentación y conocer el comportamiento de la deserción desde diferentes ópticas, tales como deserción por cohorte, según el nivel de formación profesional del responsable del estudiante, si la institución es oficial o no, el ingreso de la familia, etc. (MEN, 2008).

No obstante, una función de supervivencia no explica el comportamiento dinámico de los estudiantes en la universidad; el tiempo de permanencia de un estudiante en la universidad no está sincronizado con su avance académico, pues al final de cada periodo cursado tiene tres posibilidades: avanzar al siguiente semestre, permanecer en el mismo semestre o retirarse de la universidad. Este es el comportamiento real, y es una de las razones por las que se usa el proceso estocástico cadenas de Markov para modelar tal fenómeno.

El fenómeno de la deserción ha sido analizado desde varias ópticas: psicológica, económica, social, institucional, académica y tecnológica (Tinto, 1993). Varios autores han encontrado que ciertas condiciones y factores están asociados con una mayor probabilidad de dejar la universidad (Pascarella & Terenzini, 1980; Tinto, 1993). Algunos de estos factores están altamente correlacionados con las características demográficas de los estudiantes, como pertenecer a un estrato socioeconómico bajo, ser mujer u hombre (Leppel, 2005), edad, origen étnico, rendimiento académico (Tinto, 1993), tener padres con un bajo nivel de educación o ser la primera generación en la universidad (es decir, ser hijo de padres que no han ido a la universidad) (Choy et al., 2000), al igual que la personalidad de los estudiantes, el nivel de motivación, la autoconfianza, los intereses académicos, las habilidades para estudiar y si trabajan o no (Green, 1998).

⁴ Índice de cobertura: se refiere al número de matriculados en el sistema de educación superior en Colombia.

Algunas condiciones dependen de las instituciones de educación superior, como el asesoramiento académico, programas de monitorías, prácticas docentes, disponibilidad de recursos financieros, condiciones de acompañamiento y de bienestar en general que favorezcan la transición del colegio a la universidad (Leppel, 2005; Cabrera et al., 1993; Pascarella & Terenzini, 1980); de igual manera, el entorno social, el respaldo y el apoyo que tienen de las familias (Nora, 2002).

En resumen, la bibliografía es rica en identificar factores que conducen al fracaso y la deserción universitaria, señalando grupos o perfiles que, debido a sus características, corren un mayor riesgo de no terminar (Pascarella, 2006; Choy et al., 2000; Perry & Cabrera, 2000). Pero pese al abrumador número de estudios sobre la deserción de los estudiantes de educación superior, este tema todavía plantea un desafío para los administradores universitarios (Berger & Braxton, 1998) y para los organismos gubernamentales de Colombia y el mundo (MEN, 2010)⁵.

Así las cosas, existe la obligación de ampliar la comprensión del problema con un enfoque diferente, teniendo en cuenta las condiciones y necesidades propias de nuestra cultura y del entorno nacional. En respuesta a esto, en el presente estudio se propone y prueba un modelo matemático con la finalidad de aportar información que ayude a encontrar una probabilidad de deserción y el tiempo de permanencia que permita identificar y valorar qué factores de riesgo contribuyen a la deserción de los estudiantes de ingeniería de la Escuela, para poder desarrollar así estrategias de retención a lo largo de su transcurso de semestre a semestre, hasta su graduación.

⁵ MEN: Ministerio de Educación Superior de la República de Colombia.

2. PLANTEAMIENTO DEL PROBLEMA

2.1. Antecedentes del problema

La deserción estudiantil en cualquier etapa académica es un problema de índole mundial, motivo por el cual muchos países han dedicado tiempo y dinero para el estudio de tal fenómeno, como se verá más adelante. Según informes del DANE⁶ y del Ministerio de Educación Nacional, con base en los datos de los alumnos matriculados en el año 2011, se puede concluir que de los estudiantes que ingresan a primaria, solamente el 42,45% concluye la educación media, y de los que ingresan a la educación superior apenas el 5,15% culmina sus estudios y se gradúa (Peña & MEN, 2006; MEN, 2008; DANE, 2011).

A causa de esto, se empezaron a desarrollar investigaciones sobre la deserción en Colombia a partir del año 2003. A nivel nacional, se han publicado cuatro estudios importantes:

- El realizado por la Universidad Nacional y el Icfes⁷, en el que se hizo una definición teórica y conceptual del fenómeno, aclarando los conceptos y términos para Colombia (UN & Icfes, 2002)⁸.
- La Universidad de Antioquia llevó a cabo un estudio en el que se encuestó a los desertores y se aplicaron técnicas estadísticas de duración; en este estudio se incluyó a los estudiantes de ingeniería y ciencias económicas (Vásquez et al., 2003).
- El efectuado por el Ministerio de Educación Nacional con el apoyo de la Universidad de los Andes, en el que se hizo un análisis estadístico de supervivencia y se desarrolló el Sistema para la Prevención y Análisis de la Deserción en las Instituciones de Educación Superior (Spadies), *software* que permite a cada institución adscrita identificar y clasificar a los estudiantes en riesgo de deserción (MEN, 2008)⁹.
- En el estudio de Castaño (2004), en el cual se usan modelos de duración con un análisis de cohortes completas para analizar el fenómeno, se calcula la probabilidad de que un estudiante deserte sujeto al tiempo que ha permanecido en la universidad. Además, permite conocer cuándo hay más riesgo de deserción asociado a ciertas características

⁶ DANE: Departamento Administrativo Nacional de Estadística.

⁷ Icfes: Instituto Colombiano para la Evaluación de la Educación.

⁸ UN: Universidad Nacional.

⁹ MEN: Ministerio de Educación Nacional.

del estudiante, y logra establecer diferencias en los riesgos para los primeros semestres, comparados con los semestres avanzados (Castaño et al., 2004).

A continuación se presenta un resumen de los estudios estadísticos sobre la deserción, realizados en el exterior, clasificados en modelos estáticos desde varios aspectos, como los individuales, los académicos, los institucionales y los socioeconómicos, diferentes de los modelos markovianos. En el siguiente diagrama se puede apreciar un esquema de estos:

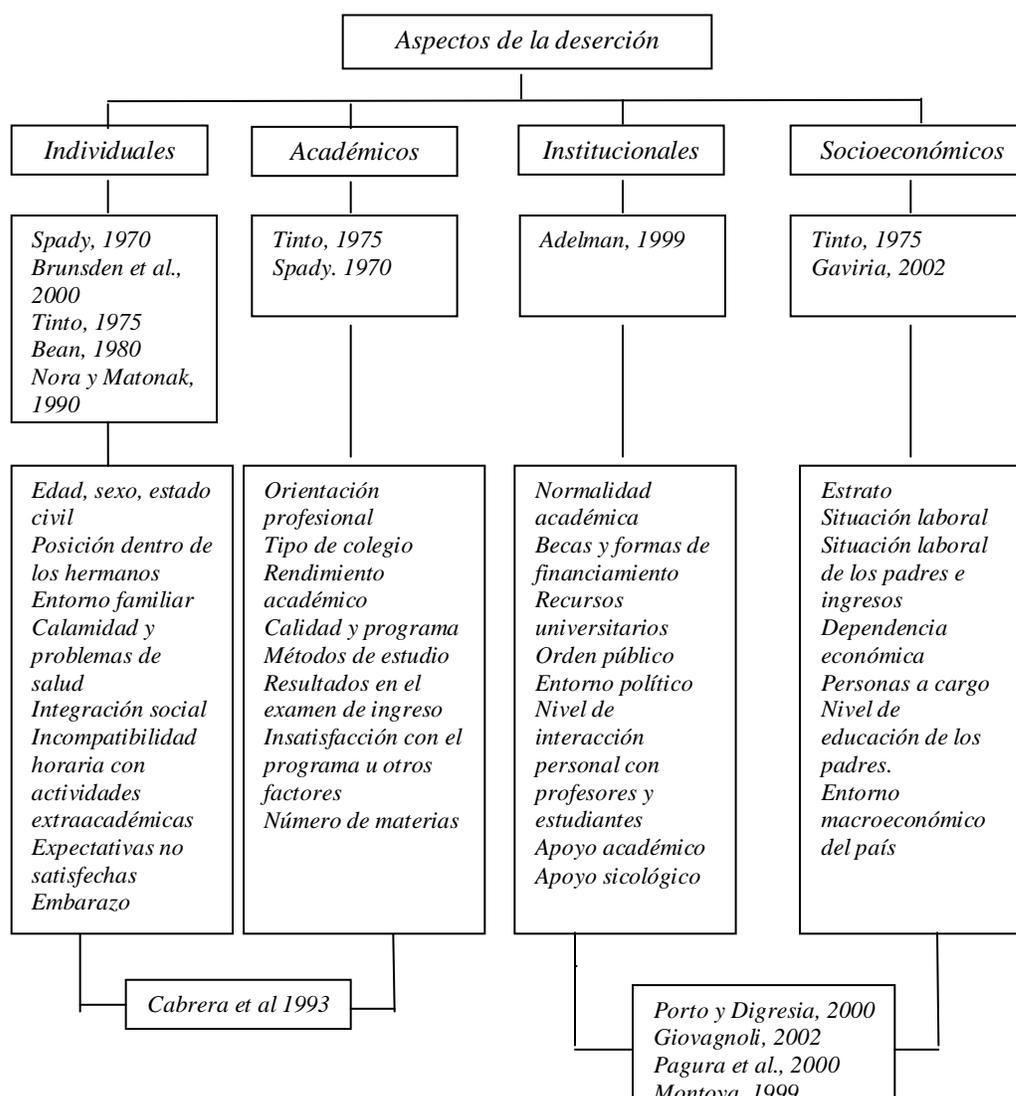


Figura 1. Estado actual de los aspectos de la deserción estudiantil (Castaño et al., 2004).

Para los estudios presentados en la gráfica de la figura 1 se usaron variables individuales y estáticas, ignorando la evolución del fenómeno a través del tiempo; además, se utilizaron modelos de ecuaciones estructurales y modelos de regresión logística (Castaño et al., 2004).

Cuando se permitió que las variables cambiaran en el tiempo, como en el estudio de Ishitani y DesJardins, en el que se desarrolló un modelo estadístico longitudinal para investigar el fenómeno de la deserción, se pudo analizar el comportamiento de las variables identificadas por ellos en el tiempo. En razón del tratamiento de las variables, sus conclusiones fueron diferentes:

- Los estudiantes de ingresos más bajos tienen mucha más probabilidad de salir de la universidad frente a los de ingresos más altos.
- Es más perjudicial para el estudiante tener ingresos bajos en el tercer y cuarto año, que en el primer año.
- Los estudiantes cuyas madres se graduaron de la universidad tienen menos probabilidades de abandonar sus estudios, siendo este efecto más fuerte en el segundo año.
- La ayuda financiera es útil para reducir la deserción y tiene diversos efectos en el tiempo (Ishitani & Desjardins, 2002).

Se presentan a continuación las investigaciones más relevantes asociadas al problema, solucionadas con el uso de cadenas de Markov.

Uno de estos estudios hace referencia al flujo de estudiantes en un programa de doctorado, en el que se definieron los estados posibles que podría tomar un estudiante y la probabilidad de cambio de un estado a otro, para predecir el número de tesis que se van a tener en algún momento. La matriz de transición de las probabilidades reveló que la tasa de deserción académica es alta para los estudiantes de los primeros semestres, en tanto que los estudiantes que han avanzado a candidatura tienen más probabilidades de permanecer y matricularse en el siguiente semestre (W. Bessent & A. Bessent, 1980).

En otro estudio destacado se aplica un modelo de Markov para predecir las probabilidades de persistencia institucional hasta la graduación. En este modelo se utilizan variables socioeconómicas que caracterizan a una minoría de los estudiantes provenientes de una base de datos de estudiantes de Houston Independent School District; uno de los resultados obtenidos fue el siguiente:

- El modelo de flujo de Markov para los estudiantes y su uso como una herramienta predictiva permite calcular un valor de riesgo de persistencia utilizando datos institucionales (A. Bessent et al., 1986).

Un estudio destacado fue Menard, donde compara el empleo de coeficientes estandarizados en general y coeficientes de regresión estandarizados logísticos en particular, y hace recomendaciones específicas sobre cuál es el mejor uso y evitar abusos de los coeficientes de regresión logísticos (Menard & Houston 2011).

Finalmente, es importante mencionar un estudio realizado por Borden en el que se usaron las cadenas de Markov para simular los cambios de curso del estudiante en las tasas de retención y graduación. Los datos se aplicaron durante un año, y permitieron registrar cómo los estudiantes cambiaban de curso o llegaban a un estado de no retorno. En el estudio se probó:

- Que un 25% de estudiantes aumentó sus calificaciones pero esto no cambió significativamente la tasa de graduación, la cual se incrementó en menos del 1%;
- Que el decrecimiento de las tasas de retención y graduación se debe en gran medida a lo inadecuado de los modelos tradicionales de educación, más que a la eficiencia institucional (Borden & Dalphin, 1998).

2.2. Justificación de la investigación

La revisión de la bibliografía existente sobre la deserción y las cifras de la realidad colombiana lleva a la conclusión de que aún se está lejos de contar con la respuesta adecuada a este fenómeno, presente en todas las instituciones de educación superior. Faltan herramientas efectivas para cuantificar los riesgos y así generar estrategias oportunas para minimizarlos, de modo que los estudiantes logren avanzar en los semestres hasta obtener su grado de manera oportuna.

Con el presente estudio se pretende identificar las características socioeconómicas, socioafectivas y académicas que pueden afectar las probabilidades de culminar con éxito un programa de Ingeniería en la Escuela. Se aplica un enfoque multivariado de análisis cuantitativo que busca investigar los efectos de los factores seleccionados sobre la

deserción académica, al igual que un modelo matemático basado en cadenas de Markov, para diseñar una herramienta de predicción.

3. FORMULACIÓN DE LA PREGUNTA DE INVESTIGACIÓN

Con base en el planteamiento presentado, se trata de responder las siguientes preguntas:

¿Cuáles son las variables que afectan la deserción académica en el programa de Ingeniería de la Escuela Colombiana de Ingeniería Julio Garavito? ¿Cuál es el tiempo que permanecen los estudiantes hasta graduarse y cuál es la probabilidad de deserción de estos, dadas unas características en su perfil como estudiante?

4. OBJETIVOS

Objetivo general

Diseñar y probar un modelo estocástico, usando cadenas de Markov para pronosticar la deserción académica de los estudiantes de ingeniería de la Escuela Colombiana de Ingeniería Julio Garavito.

Objetivos específicos

1. Determinar si se presenta mayor deserción en los primeros semestres de ingeniería.
2. Precisar cuál es el tiempo de permanencia de un estudiante desde que empieza su carrera hasta que se gradúa.
3. Investigar cuáles son las variables que afectan la probabilidad de deserción de un estudiante.
4. Analizar el comportamiento del fenómeno de la deserción con base en el cambio de los posibles valores de las variables independientes, dados unos perfiles de los estudiantes.
5. Proponer recomendaciones para resolver el problema de la deserción en ingeniería en la Escuela, con base en los resultados obtenidos al aplicar el modelo matemático diseñado.

5. ESTADO DEL ARTE

Según Leppel, los porcentajes de deserción varían entre diferentes países, pero lo que es invariante es la correlación entre la deserción de los estudios y la reducción del mercado laboral para el individuo. Las investigaciones a nivel nacional e internacional indican consistentemente que a mayor deserción hay mayor probabilidad de desempleo, menores ingresos y, en general, una mínima riqueza acumulada a lo largo de la vida. También son menores la salud física y mental, la participación en las actividades cívicas, así como el aumento en las tasas de criminalidad (Leppel, 2005).

5.1. Deserción

El término deserción se usa para referirse a los jóvenes que dejan la universidad antes de graduarse (Lamb et al., 2011).

Hay dos tipos de deserción: la exclusión académica y la deserción voluntaria (Tinto & Cullen, 1975). La deserción puede tener diferentes ópticas, según la persona que la observe; por ejemplo, si la analiza un funcionario de una universidad la puede ver como un fracaso de los estudiantes, mientras que el alumno la puede ver como un éxito porque fue capaz de decidir que eso no le gustaba y cambiar. El hecho de que abandone los estudios en el primer semestre no es un fracaso para él, ya que es tan valiente que decide cambiar el rumbo de su destino e interesarse en otra carrera que realmente le llene las expectativas (Tinto, 1975); por esto, Tinto dice que la deserción es un comportamiento individual.

Tinto destaca el proceso de la deserción en tres etapas: una inicial, en la que el estudiante trae unas expectativas y unas ideas sobre cómo es la universidad, estas pueden estar equivocadas y se desilusionan en el primer semestre, de ahí que este semestre sea crítico para todos los programas (Tinto, 2010). Para prevenir esto, universidades como los Andes, la Escuela, el Rosario, reúnen a los padres de los estudiantes admitidos, les muestran las bondades de los programas y del campus, los servicios de Bienestar Universitario, las estadísticas del programa, los campos de acción de los egresados, los convenios internacionales, etc., para ayudarles a decidir sobre dónde matricular finalmente a sus hijos.

Aquí se aprecian dos cambios en la historia de los procesos de admisión: uno es el acompañamiento más cercano hoy en día por parte de los padres a los admitidos a primer semestre y otro es que estos estudiantes están llegando mucho más jóvenes a la universidad que en años anteriores (MEN, 2010)¹⁰.

Una segunda etapa es el primer semestre como tal. La difícil adaptación por parte de los estudiantes a la cultura de la universidad, el tema de los trabajos, las entregas de proyectos en grupo, el manejo del tiempo, la organización, los métodos de estudio, la adaptación al medio social y al ambiente de la institución forman parte de un periodo de transición que puede resultar peor aún para las minorías (Tinto, 2010). La tercera etapa se presenta en los primeros semestres de las carreras profesionales, donde las ciencias básicas son la barrera para lograr avanzar y, finalmente, graduarse (Tinto, 2010; Universidad Nacional de Colombia & Icfes, 2002).

Se han detectado variables como los ingresos familiares, el promedio de notas del colegio, las relaciones sociales, que influyen en la deserción de los primeros semestres, así como las políticas institucionales en cuanto a los procesos de inicio, sistemas de comunicación y procesos de admisión (Berger & Braxton, 1998). Se encontró también que aspectos como el apoyo de la familia y el contacto con los amigos son claves para evitar la deserción (Nora, 2002).

En varios estudios (Spady, 1970; Cabrera et al., 1993; Castaño et al., 2004; Nora, 2002; Berger & Braxton, 1998, y Donoso & Schiefelbein, 2007) se propone clasificar las perspectivas de la deserción, las cuales cambian dependiendo de las variables –personales, familiares o institucionales–; estas son: perspectiva psicológica, sociológica, económica, organizacional e interaccionista.

Desde la perspectiva psicológica, lo que influye fundamentalmente en la deserción es la personalidad del estudiante, ya que existen rasgos que diferencian a las personas que logran

¹⁰ MEN: Ministerio de Educación Nacional.

el grado de las que no. La decisión cambia por las conductas previas, por las creencias y por las normas (Ajzen & Fishbein, 1977).

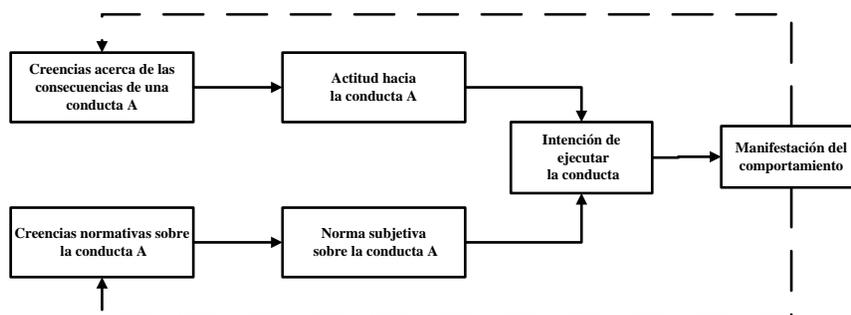


Figura 2 Modelo de Fishbein y Ajzen (Ajzen & Fishbein, 1977).

Desde la perspectiva sociológica, Ethington amplía el concepto –con base en los estudios anteriores– e identifica que el apoyo y el estímulo de la familia favorecen la autoestima y las aspiraciones de los estudiantes, quienes inciden a su vez en la percepción de los problemas y la forma de enfrentarlos, lo cual finalmente redundando en una menor deserción (Ethington, 1990).

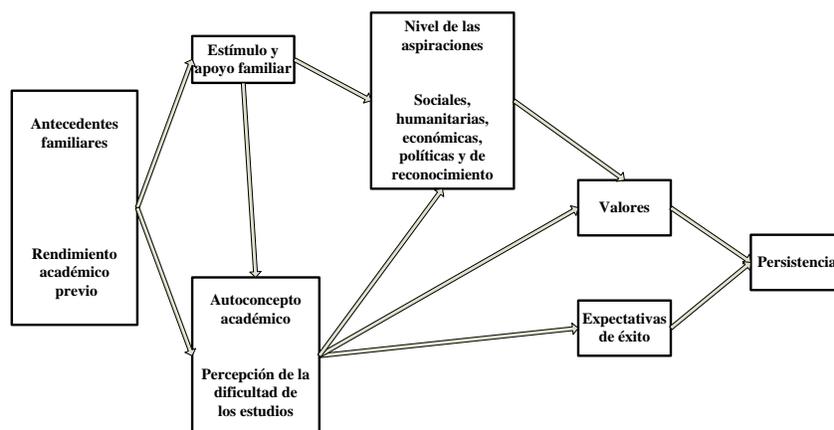


Figura 3 Modelo de Ethington (Ethington, 1990).

Por su parte, Sapdy y Harder sostienen que la familia también influye en el potencial académico, así como en las normas y su congruencia, afectando no sólo el rendimiento académico sino el apoyo de los compañeros y las relaciones sociales con su entorno, lo que además está directamente ligado con la satisfacción del estudiante y su compromiso con la institución (Sapdy, 1971; Hader, 2011).

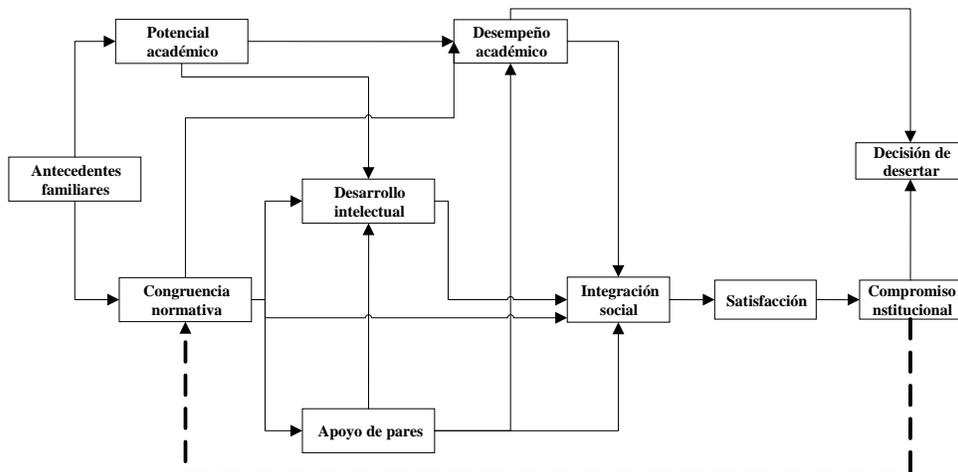


Figura 4 Modelo sociológico de deserción Spady (Spady, 1970).

Desde la perspectiva económica, se tienen dos clasificaciones: el modelo costo-beneficio y el apoyo mediante subsidios. Se concluye que si los beneficios sociales y económicos que perciben los estudiantes son menores comparados con el esfuerzo y el trabajo que se requieren para desarrollar la carrera, las personas terminan por abandonar los estudios. Igualmente influye si el estudiante tiene la percepción de capacidad económica para poder cubrir sus estudios, para lo cual las instituciones ofrecen subsidios, becas, créditos a largo plazo y con tasas de interés muy bajas, a los cuales pueden acceder dichos estudiantes (Cabrera et al., 1993; Castaño et al., 2004; Nora, 2002).

Otros estudios, como el de Porto, indican que un factor determinante en la deserción es el nivel académico de los padres, pues a mejor formación de estos, menor la deserción de sus hijos; el nivel más alto de educación implica un mejor entorno socioeconómico para el estudiante (Porto & Gresia, 2001).

Por otra parte, desde la perspectiva organizacional se ha encontrado que a mayor tamaño de las instituciones, mejores herramientas de apoyo se les pueden brindar a los estudiantes, como monitorías, tutorías, seguimientos, instalaciones deportivas, servicios de psicología, servicios de bienestar y consecución de empleos para los egresados, con lo que se disminuyen los indicadores de deserción (Berger & Braxton, 1998; John et al., 2000; Tinto, 2007).

En cuanto a la perspectiva interaccionista, se explica como la construcción de una integración social y académica del estudiante con la institución, que se logra con la adaptación del estudiante a la vida universitaria y a la exigencia académica. En este modelo, basado en el de Spady pero que lo complementa, influyen los antecedentes familiares, las destrezas y habilidades del estudiante, y las intenciones; además, se separa claramente el sistema social del sistema académico, pero a su vez las consecuencias de los dos se integran para afectar la decisión de la deserción, como se puede observar en la siguiente figura (Tinto, 1993).

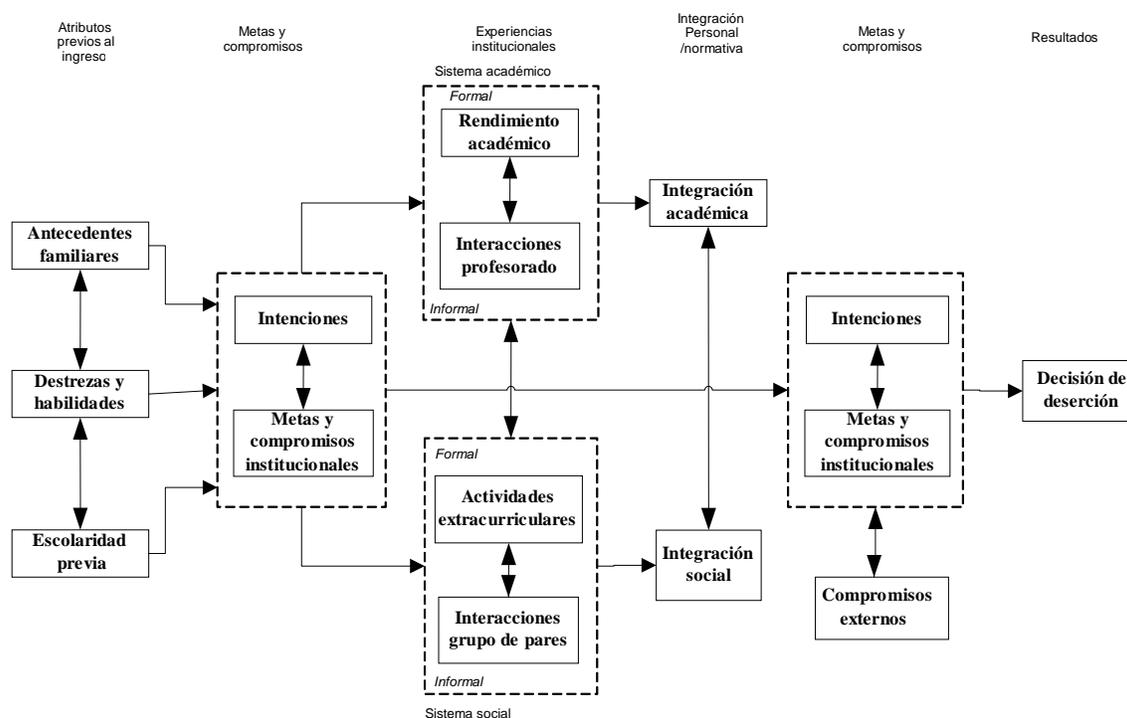


Figura 5 Modelo de integración de Tinto (Tinto, 1993).

Todas las definiciones y los sistemas comparten un entendimiento similar de lo que es la deserción. Sin embargo, no ocurre lo mismo cuando se quiere medir la deserción o cuantificarla, ya que hay muy pocas medidas comunes; incluso en cada país no hay claridad en cómo medir la deserción (Lamb et al., 2011).

La educación superior está dividida en niveles, localizados en diferentes clases de universidades y que llevan a varios tipos de calificaciones (profesionales, técnicas o vocacionales). La diversidad de programas y calificaciones produce conflictos en temas de

equivalencia en comparaciones entre países. Las mediciones de la deserción o de la graduación no son necesariamente complementarias u opuestas exactas, ya que es posible que algunos estudiantes que salen de una institución completen luego su formación en otra (MEN, 2008).

Así mismo, es factible que varios estudiantes permanezcan en la institución hasta el final y en ciertos casos cuenten como terminados; sin embargo, estos estudiantes dejan de cumplir con sus requisitos de grado y no se gradúan. Igualmente, en algunos sistemas existe también un nivel de flexibilidad que permite a estudiantes cambiar de programas o estar en programas sin restricción de edad; en este caso, las tasas de deserción y graduación variarán, dependiendo del momento o el rango de edad que se decida medir (Lamb et al., 2011).

5.2. Cadenas de Markov

Un proceso estocástico $\mathbf{X}=\{X_t, t \in T\}$ es una colección de variables aleatorias. Para cada t en el conjunto de índices T , X_t es una variable aleatoria. Usualmente se interpreta t como el tiempo y $X_t =j$ es llamada el estado del proceso en el tiempo t . El índice t puede ser un conjunto contable, en cuyo caso el proceso estocástico se denomina proceso estocástico de tiempo discreto; en caso contrario, es llamado de tiempo continuo (Ross, 1996).

Se dice que un proceso estocástico tiene incrementos independientes si para todo $t_0 < t_1 < t_2 \dots < t_n$ las variables aleatorias $X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$ son independientes. Y el proceso posee incrementos estacionarios si $X_{t_{m+n}} - X_{t_m}$ tiene la misma distribución para todo m (Ross, 1996).

Una cadena de Markov es un proceso estocástico de tiempo discreto, que satisface la propiedad markoviana si para cualquier n la distribución condicional de X_{n+1} , dadas X_0, X_1, \dots, X_n , es independiente de X_0, X_1, \dots, X_n ,

$$P\{X_{n+1}=j \mid X_0=i_0, X_1=i_1, \dots, X_n=i\} = P\{X_{n+1}=j \mid X_n=i\} = p_{ij}$$

Es decir, la probabilidad de estar en el estado $n+1$ sólo depende del estado inmediatamente anterior del sistema,

$$P(X_{n+1}=j | X_n = i) = p_{ij}$$

Si el sistema se mueve del estado i al estado j durante el siguiente periodo, se dice que ocurrió una transición de i a j . El conjunto de las p_{ij} se denomina probabilidades de transición de la cadena de Markov (Winston, 2008).

Las cadenas de Markov se pueden representar como una matriz cuadrada de $N+1 \times N+1$ de probabilidades, llamada matriz de transición, donde las filas corresponden al estado de inicio y las columnas al estado de llegada:

$$P = \begin{bmatrix} p_{00} & \dots & p_{0N} \\ \vdots & \ddots & \vdots \\ p_{N0} & \dots & p_{NN} \end{bmatrix}$$

Dado que el estado en el tiempo n es i , el proceso en alguna parte debe estar en el tiempo $n+1$. Esto significa que para cada i ,

$$\sum_{j=1}^N (PX_{n+1} = j | P(X_n = i)) = 1$$

$$\sum_{j=1}^N P_{ij} = 1 \quad 1 \leq i \leq N.$$

Cada elemento de la matriz P debe ser no negativo y la suma de los elementos de cada renglón debe ser igual a 1.

Las matrices de transición se pueden representar mediante gráficas, donde cada nodo representa un estado y un arco (i, j) representa la probabilidad de transición p_{ij} . (Winston, 2008).

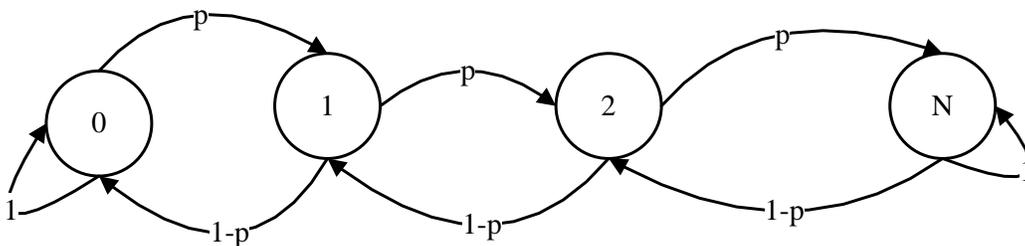


Figura 6. Ejemplo de una notación gráfica de una matriz de transición.

Los estados de las cadenas de Markov se clasifican en:

Estado alcanzable. Dados dos estados i y j , una trayectoria de i a j es una secuencia de transiciones que comienza en i y termina en j , tal que cada transición en la secuencia tiene una probabilidad positiva de ocurrir. Un estado j es alcanzable desde el estado i si hay una trayectoria que conduzca de i a j .

Estados que se comunican. Se dice que dos estados i y j se comunican si j es alcanzable desde i , e i es alcanzable desde j .

Conjunto cerrado. Un conjunto de estados S de una cadena de Markov es cerrado si ningún estado fuera de S es alcanzable desde algún estado en S .

Estado absorbente. Es un estado en el que $p_{ii} = 1$ $p_{ij} = 0$. Siempre que se entra en un estado absorbente, no se sale de él; un estado absorbente es un conjunto cerrado que contiene sólo un estado.

Estado transitorio. Un estado i es transitorio si existe un estado j que es alcanzable desde i , pero el estado i no es alcanzable desde el estado j .

Estado recurrente. Es un estado que no es transitorio.

Estado periódico. Un estado i es periódico con periodo $k > 1$ si k es el número más pequeño tal que las trayectorias que conducen al estado i de regreso al estado i tienen una longitud que es un múltiplo de k . Si un estado recurrente no es periódico, se conoce como aperiódico (Winston, 2008).

Cadenas absorbentes. Una cadena de Markov absorbente es una cadena en la cual algunos de sus estados son absorbentes y el resto son transitorios. Estas cadenas absorbentes tienen unas propiedades importantes que se explican a continuación.

En las cadenas absorbentes se reorganizan los estados de la siguiente manera: primero los estados transitorios, luego los estados absorbentes. Si hay $s - m$ estados transitorios (t_1, t_2, \dots, t_{s-m}) y m estados absorbentes (a_1, a_2, \dots, a_m). Se escribe la matriz de probabilidades de transición P así (Winston, 2008):

$$\begin{array}{c} s - m \text{ renglones} \\ m \text{ renglones} \end{array} P = \begin{array}{cc} \begin{array}{c} s - m \\ \text{Columnas} \end{array} & \begin{array}{c} m \\ \text{Columna} \end{array} \\ \left[\left\{ \begin{array}{c|c} Q & R \\ \hline 0 & I \end{array} \right\} \right] \end{array}$$

Si una cadena comienza en un estado transitorio, antes de que llegue a un estado absorbente, ¿cuántos periodos se espera pasar en un determinado estado transitorio antes

que tenga lugar la absorción? Si en el presente se está en un estado transitorio t_j , el número esperado de periodos que se pasará en un estado transitorio t_j , antes de la absorción, es el ij -ésimo elemento de la matriz $(I - Q)^{-1}$ (Winston, 2008).

Si una cadena empieza en un estado transitorio, ¿cuál es la probabilidad de que se termine en cada estado absorbente? Si en el presente estamos en un estado transitorio t_j , la probabilidad de que finalmente seamos absorbidos en un estado absorbente a_j es el ij -ésimo elemento de la matriz $(I - Q)^{-1} * R$ (Winston, 2008).

6. METODOLOGÍA DE LA INVESTIGACIÓN

6.1. Población y muestra

Se contó con la información académica almacenada en la base de datos de la Secretaría de la Escuela, la cual tiene datos desde el año 2000 hasta el año 2011. No obstante, se requirió información socioeconómica que se hallaba en poder del área financiera, que también se solicitó para todos estos estudiantes, complementando así la información de base.

Al hacer un análisis de los datos se observa que sólo a partir del año 2003 se tienen los datos de los créditos académicos, los cuales son claves para los cálculos requeridos, razón por la cual se decidió no tomar las cohortes anteriores a este año; igualmente, existen estudiantes que vienen de transferencia de otras universidades, pero es apenas un 0,1% del total, lo cual no es significativo para el estudio y por eso no se tiene en cuenta.

La muestra que se seleccionó son ocho cohortes completas de los estudiantes de ingeniería de la Escuela Colombiana de Ingeniería Julio Garavito, que se identifican a partir del primer semestre del año 2003-(1 y 2), 2004-(1 y 2), 2005-(1 y 2) y 2006-(1 y 2), hasta el segundo semestre del año 2010-2. Esta información es de 3734 estudiantes, que equivale a una base de 143.401 registros, con la información socioafectiva, socioeconómica, demográfica y académica que se tiene almacenada en la institución. Los datos iniciales son suministrados en archivos en planos.

6.2. Diseño del estudio

Este es un estudio de caso cuantitativo y el diseño de la investigación es no experimental, longitudinal de evolución de grupo o cohorte, porque se toman los datos de cohortes completas y se analiza su comportamiento a lo largo del tiempo. No obstante, se hacen unas simulaciones con unos perfiles de estudiantes con características especiales para ver su comportamiento con el modelo estocástico que se diseñó, con el que se predicen el tiempo de permanencia y las probabilidades de graduarse y de retirarse de cualquiera de los programas de Ingeniería en la Escuela Colombiana de Ingeniería Julio Garavito.

La metodología que se usó en la investigación se sustenta en lo propuesto en el “Modelo de flujos de educación estudiantil, usando cadenas de Markov” (W. Bessent & A. Bessent, 1980). Es así como un estudiante va fluyendo a través de los semestres uno a uno, hasta alcanzar el grado. Una vez que ingresa a primer semestre, tiene tres posibilidades: la primera es que luego de cursar el semestre se retire de la institución, lo que para efectos de código interno se denomina SsR y se considera como una deserción; en el modelo de Markov es un estado absorbente, porque luego de que llega a este no se puede devolver. Una segunda posibilidad es que el estudiante pierda el semestre, código interno SsP, y decida repetirlo; entonces sigue en el programa. Por último, existe la posibilidad de que el estudiante apruebe su semestre, código SsA, y avance sucesivamente hasta que se gradúe; el graduarse también es un estado absorbente (W. Bessent & A. Bessent, 1980). Esto se observa así en una representación gráfica:

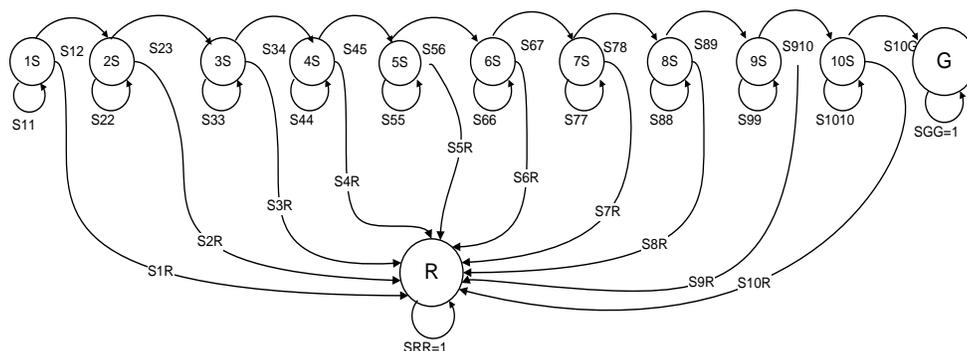


Figura 7 Cadena de Markov de los estudiantes semestre a semestre. Elaboración propia.

Cada nodo representa un semestre académico, entonces 1S es primer semestre, 2S es segundo semestre y así sucesivamente; están los diez nodos de cada uno de los semestres y

dos nodos especiales: cuando el estudiante se retira de la universidad pasa a un estado R, que significa que se va y no vuelve más, o cuando se gradúa, que pasa a un estado G, que se supone que tampoco vuelve más al sistema. Estos dos últimos son estados absorbentes, porque después que entra allí, el estudiante no sale. La probabilidad de estos dos estados absorbentes es 1.

Por otra parte, al estar en el estado 1S tiene tres probabilidades: una es S12, que es la probabilidad de avanzar del estado 1S a 2S, esto es, de avanzar de primer semestre a segundo semestre. La otra probabilidad es S1R, que significa que puede pasar al estado R. Por último, puede que permanezca en el estado 1S, lo cual quiere decir que tiene la probabilidad S11. Igual para los diez semestres en los que puede transitar antes de graduarse, o antes de retirarse.

Esta gráfica se puede representar en una matriz de transiciones, así:

	1S	2S	3S	4S	5S	6S	7S	8S	9S	10S	Gradúan	Retiran
1S	S11	S12	0	0	0	0	0	0	0	0	0	S1R
2S	0	S22	S23	0	0	0	0	0	0	0	0	S2R
3S	0	0	S33	S34	0	0	0	0	0	0	0	S3R
4S	0	0	0	S44	S45	0	0	0	0	0	0	S4R
5S	0	0	0	0	S55	S56	0	0	0	0	0	S5R
6S	0	0	0	0	0	S66	S67	0	0	0	0	S6R
7S	0	0	0	0	0	0	S77	S78	0	0	0	S7R
8S	0	0	0	0	0	0	0	S88	S89	0	0	S8R
9S	0	0	0	0	0	0	0	0	S99	S910	0	S9R
10S	0	0	0	0	0	0	0	0	0	S1010	S10G	S10R
Gradúan	0	0	0	0	0	0	0	0	0	0	1	0
Retiran	0	0	0	0	0	0	0	0	0	0	0	1

Figura 8 Matriz de transición de semestres. Elaboración propia.

Las filas representan los estados iniciales que puede tener el estudiante y las columnas el estado siguiente, los ceros representan que no es posible ese cambio de estado y los valores son las 30 probabilidades que se tienen; estas probabilidades deben sumar 1 por cada fila.

6.3. Variable Y_n

Para modelar el proceso markoviano se definió la variable Y_n como el promedio ponderado de las notas y los créditos de las asignaturas cursadas en un semestre n . Esta variable es categórica ordenada y toma tres valores: 1 si el estudiante se retira en ese semestre, 2 si el estudiante pierde el semestre y 3 si el estudiante pasa al siguiente semestre.

Cuando un alumno finaliza un periodo de estudio, obtiene unas notas de las asignaturas cursadas; con base en estas y sus respectivos créditos académicos, se calcula el promedio ponderado del semestre: si el promedio ponderado da menor que 3,0, el estudiante pierde el semestre y a la variable dependiente se le asigna 2. Si, por el contrario, el promedio es mayor o igual que 3,0, se le asigna 3 a la variable dependiente. Si el estudiante se retira, se le coloca 1 a la variable dependiente, en el último semestre que el estudiante haya cursado.

El problema al cual se enfrentó era cómo calcular las probabilidades de cambio de un estado a otro. Una primera aproximación fue el análisis de frecuencias que arrojó la matriz de frecuencias inicial, de cómo se distribuye la variable dependiente en los diez semestres académicos.

	1 Sem	2 Sem	3 Sem	4 Sem	5 Sem	6 Sem	7 Sem	8 Sem	9 Sem	10 Sem	Gradúan	Retiran	Total
1 Sem	708	3093										568	4369
2 Sem		521	2507									464	3492
3 Sem			363	2240								272	2875
4 Sem				325	2065							193	2583
5 Sem					263	1935						136	2334
6 Sem						223	1813					147	2183
7 Sem							173	1681				135	1989
8 Sem								139	1470			107	1716
9 Sem									228	1037		67	1332
10 Sem										192	1136	59	1387

Tabla 1 Frecuencias iniciales de la variable dependiente.

Estas frecuencias se representan en probabilidades en la siguiente cadena de Markov:

	1 Sem	2 Sem	3 Sem	4 Sem	5 Sem	6 Sem	7 Sem	8 Sem	9 Sem	10 Sem	Gradúan	Retiran	Suma
1 Sem	0,16205	0,70794	0	0	0	0	0	0	0	0	0	0,13001	1
2 Sem	0	0,1492	0,71793	0	0	0	0	0	0	0	0	0,13288	1
3 Sem	0	0	0,12626	0,77913	0	0	0	0	0	0	0	0,09461	1
4 Sem	0	0	0	0,12582	0,79946	0	0	0	0	0	0	0,07472	1
5 Sem	0	0	0	0	0,11268	0,82905	0	0	0	0	0	0,05827	1
6 Sem	0	0	0	0	0	0,10215	0,83051	0	0	0	0	0,06734	1
7 Sem	0	0	0	0	0	0	0,08698	0,84515	0	0	0	0,06787	1
8 Sem	0	0	0	0	0	0	0	0,081	0,85664	0	0	0,06235	1
9 Sem	0	0	0	0	0	0	0	0	0,17117	0,77853	0	0,05030	1
10 Sem	0	0	0	0	0	0	0	0	0	0,13843	0,81903	0,04254	1
Gradúan	0	0	0	0	0	0	0	0	0	0	1	0	1
Retiran	0	0	0	0	0	0	0	0	0	0	0	1	1

Tabla 2 Matriz de probabilidades iniciales de la variable dependiente.

Esta primera aproximación no es útil porque no cumple las propiedades requeridas para que el proceso sea considerado markoviano; igualmente, el comportamiento de los datos no cumple las condiciones de homogeneidad y normalidad. Además, algunas variables importantes son categóricas y la variable de estudio es categórica y ordinal; por lo anterior, se decidió usar la ecuación de regresión logística multinomial para calcular las probabilidades de transición en cada uno de los semestres académicos de los programas de Ingeniería de la Escuela (A. W. Astin & Denson, 2009; Hair et al., 1999; Ishitani & Desjardins, 2002; Liao, 1994; A. Bessent et al., 1986).

El modelo utilizado para el cálculo de las probabilidades fue (Liao 1994)¹¹, el detalle de la deducción completa se puede observar en el anexo 7.

$$y^* = \sum_{k=1}^K \beta_k x_k + \epsilon$$

$$y = 1 \text{ si } y^* \leq \mu_1 (= 0),$$

$$= 2 \text{ si } \mu_1 \leq y^* \leq \mu_2,$$

$$= 3 \text{ si } \mu_2 \leq y^* \leq \mu_3,$$

$$\vdots$$

$$\vdots$$

$$\vdots$$

$$= J \text{ si } \mu_{j-1} \leq y^* ,$$

Basado en el modelo de respuesta general ordinal de j categorías para las k variables explicativas y $i = 1, 2, \dots, n$ individuos, el modelo es:

$$\text{Prob}(y \leq j|x) = \alpha_j + \beta \cdot x_i$$

Donde β es un vector de $k \times 1$, del intercepto de los parámetros, x_i es un vector de $1 \times k$ de las variables del i -ésimo individuo, α_j es el intercepto del j -ésimo valor categórico. Para tres categorías, la probabilidad de la i -ésima es

$$\text{Prob}(y_j = 1|x_i) = F(\alpha_1 + \beta \cdot x_i)$$

$$\text{Prob}(y_j = 2|x_i) = F(\alpha_2 + \beta \cdot x_i)$$

¹¹ Este modelo estadístico fué tomado del libro *Interpreting Probability Models Logit, Probit, and Other Generalized Linear Models* de Tim Futing Liao.

$$Prob(y_j = 3|x_i) = 1 - F(\alpha_2 + \beta \cdot x_i)$$

Donde F es la función acumulativa Logit,

$$F(x) = \frac{1}{1 + e^{-(\alpha_i + \sum \beta \cdot x)}}$$

Ya que la probabilidad de transición es el objeto de este estudio, se considera más apropiado trabajar con los betas en vez de usar las proporciones de probabilidades; es decir, se utilizarán los betas y no los *odd ratios* del modelo, lo cual es una práctica común de los modelos Logit (DeMaris, 1993; Liao, 1994). A continuación se definen las variables que se usaron en el modelo.

6.4. Variables independientes

Las variables independientes se clasificaron en cuatro grupos: socioafectivas, socioeconómicas, académicas y demográficas, las cuales se presentan en la siguiente tabla:

Categoría	Nombre variable	Tipo	Valores
Demográficas	Sexo	Ordinal	1= Mujer 2= Hombre
Demográficas	Edad	Continua	Entre 15 y 32
Socioafectivas	Vive actualmente	Ordinal	1= vive solo 2= vive con los padres 3= vive con la madre 4= vive con el padre 5= otro
Socioafectivas	Tipo de vivienda	Ordinal	1= propia y pagada 2= propia y deuda 3= arrendada 4= familiar 5= otra
Socioafectivas	Responsable	Ordinal	0= ninguno 1= madre 2= padre 3= otro
Socioafectivas	Nivel de educación del responsable	Ordinal	0= ninguno 1= primaria 2= secundaria 3= técnica 4= universitaria o más
Socioeconómicas	Estrato	Ordinal	1, 2, 3, 4, 5 y 6
Socioeconómicas	Valor matrícula	Continua	Entre 2.475.000 y 6.670.000
Socioeconómicas	Total ingresos	Continua	Entre 480.000 y

Categoría	Nombre variable	Tipo	Valores
			60.000.000
Socioeconómicas	Naturaleza colegio	Ordinal	1= oficial 2= no oficial 3= otro
Socioeconómicas	Calendario colegio	Ordinal	1= A 2= B 3= otro
Académicas	Icfes Matemáticas	Continua	Entre 20 y 90
Académicas	Icfes Física	Continua	Entre 20 y 90
Académicas	Icfes Lenguaje	Continua	Entre 20 y 90
Académicas	AGEO – Análisis geométrico	Continua	Entre 0 y 5,0
Académicas	PREC – Precálculo	Continua	Entre 0 y 5,0
Académicas	FFIS – Fundamentos de Física	Continua	Entre 0 y 5,0
Académicas	ALLI – Álgebra lineal	Continua	Entre 0 y 5,0
Académicas	CALD – Cálculo diferencial	Continua	Entre 0 y 5,0
Académicas	CALI – Cálculo integral	Continua	Entre 0 y 5,0
Académicas	CALV – Cálculo vectorial	Continua	Entre 0 y 5,0
Académicas	ECDI – Ecuaciones diferenciales	Continua	Entre 0 y 5,0
Académicas	QUIM – Química	Continua	Entre 0 y 5,0
Académicas	BIOL – Biología	Continua	Entre 0 y 5,0
Académicas	FIS1 – Física 1	Continua	Entre 0 y 5,0
Académicas	FIS2 – Física 2	Continua	Entre 0 y 5,0
Académicas	FIS3 – Física 3	Continua	Entre 0 y 5,0

Tabla 3 Variables independientes del modelo. Elaboración propia.

Variables administrativas

Las siguientes se clasifican como variables administrativas porque sirven para la identificación y clasificación de los registros, pero no se incluyen dentro del análisis en sí de los datos.

Categoría	Nombre variable	Tipo	Valores
Administrativas	Id estudiante	Ordinal	De 2060415 a 3071582
Administrativas	Programa	Ordinal	1= industrial 2= electrónica 3= civil 4= eléctrica 5= sistemas
Administrativas	Código asignatura	Texto	Nemónicos de los cursos.
Administrativas	Nota	Continua	Entre 0 y 5,0
Administrativas	Estado	Ordinal	1= terminó estudios 2= excluido 3= no pago 4= activo
Administrativas	Año ingreso	Continuo	Entre 2003 y 2006
Administrativas	Periodo de ingreso	Continuo	Entre 1 y 2
Administrativas	Año académico	Continuo	Entre 2003 y 2010
Administrativas	Periodo académico	Continuo	1= primer periodo del año 2= segundo periodo del año 3= periodo intersemestral
Administrativas	Semestre	Ordinal	De 1 a 10
Administrativas	Variable dependiente	Ordinal	1= se retira 2= pierde 3= avanza

Tabla 4 Variables administrativas del modelo. Elaboración propia.

6.5. Procedimientos

Una vez se obtuvo la base de datos se procedió a hacer los procedimientos de reducción, de recodificación, de integración y de limpieza respectiva de los mismos. Se hizo un programa para identificar la variable dependiente por cada uno de los periodos que cursaba el estudiante.

Posteriormente se hizo otro programa para saber a qué semestre correspondía el periodo analizado, con base en las notas y los créditos de las notas. Según el número de créditos aprobados se sabe en qué semestre académico se encuentra el estudiante, lo cual es muy importante porque por ejemplo una persona puede cursar cuatro periodos pero estar en segundo semestre académico.

Cuando se tuvo la base correcta, se separó por semestres para correr el proceso de regresión logística. Entonces se procedió a analizar cuáles son las variables significativas y se corrieron varios modelos hasta obtener el más adecuado. Se hizo un análisis de su validez interna para asegurar el óptimo válido, con base en el criterio de aleatoriedad proporcional y el criterio de aleatoriedad máxima.

Con los betas disponibles para cada semestre, se usó un programa que se diseñó para calcular las matrices de transición de probabilidad de los estados de los semestres, con el fin de conocer el comportamiento a través del sistema. Con base en estas matrices fundamentales se calcularon las probabilidades de retirarse de los estudiantes dadas unas condiciones en las variables independientes, se analizaron los resultados y se presentaron las conclusiones de los hallazgos encontrados.

7. RESULTADOS

En esta sección se presentan los resultados de los análisis de datos. En primer lugar se muestran el manejo y la administración de los datos, proceso que contempla la reducción, limpieza, recodificación e integración de los datos, y después el análisis de los datos perdidos y su tratamiento. En segundo término se presenta la estadística descriptiva de los datos, en tercera instancia se describen los programas para cálculo de la variable dependiente, en cuarto lugar se hace la descripción del proceso y se muestran los resultados de la regresión logística multinomial aplicada a cada uno de los diez semestres, y finalmente se presenta la aplicación de las matrices de transición de la cadena de Markov aplicada a perfiles planteados, para dar respuesta a las preguntas del estudio y sus resultados.

7.1. Reducción y limpieza de los datos

- Para los procedimientos de reducción y limpieza, así como el de recodificación e integración de los datos, se usa la estructura de datos original suministrada por la Secretaría General de la Escuela, que se presenta en detalle en el anexo 4.

- El detalle de las acciones realizadas en el proceso de reducción y limpieza de datos se presenta en el anexo 5.

7.2. Recodificación de los datos

El detalle de las acciones realizadas en el proceso de reducción y limpieza de datos se presenta en el anexo 6.

7.3. Integración de los datos

- Se une la información financiera de los estudiantes con la base que tienen las notas y las asignaturas de cada estudiante.
- Se reducen los responsables a uno solo, porque estaban distribuidos en tres; lo que ocurrió fue que sólo estaba el primero casi siempre y los demás quedaban en null, por lo cual causaba problemas por los valores perdidos, lo que al final no tenía sentido.
- Igual se unen los ingresos de la familia con la renta reportada, porque casi siempre que había una la otra estaba en null, causando problemas con los datos perdidos; aquí se tuvo en cuenta que la renta es anual y el ingreso es mensual, razón por la cual se pasó todo a mensualidades.

7.4. Estadística descriptiva de los datos

En esta sección se presenta la descripción estadística de los datos utilizados. La clasificación de los estudiantes por cohorte se puede observar en la siguiente tabla:

	Año ingreso				Total
	2003	2004	2005	2006	
	Recuento	Recuento	Recuento	Recuento	
Estudiantes	1077	956	788	913	3734
Id	44320	37512	30200	31369	143401

Tabla 5 Descripción de los estudiantes por cohorte. Elaboración propia.

A continuación se presentan las frecuencias de las variables independientes categóricas.

		Recuento	% del N de la capa
Sexo	Mujer	48230	33,6%
	Hombre	95171	66,4%
Estrato	1	2220	1,5%
	2	28576	19,9%
	3	77200	53,8%
	4	27911	19,5%
	5	6432	4,5%
	6	1062	,7%
Vive actualmente	Vive solo	153	,1%
	Vive con padres	108852	75,9%
	Vive con madre	23368	16,3%
	Vive con padre	1437	1,0%
	Otros	9591	6,7%
Responsable	Madre	128488	89,6%
	Padre	6234	4,3%
	Otro	8679	6,1%
Nivel educ. Resp	Ninguna	6441	4,5%
	Primaria	11719	8,2%
	Secundaria	60550	42,2%
	Técnica	22706	15,8%
	Universitaria o más.	41985	29,3%
Tipo Vivienda	Propia y Pagada	82152	57,3%
	Propia y Deuda	24834	17,3%
	Arrendada	25065	17,5%
	Familiar	663	,5%
	Otra	10687	7,5%
Naturaleza	OFICIAL	36664	25,6%
	NO_OFICIAL	106737	74,4%
Calendario	A	137149	95,6%
	B	4040	2,8%
	F	2212	1,5%
Programa	INDUSTRIAL	52482	36,6%
	ELECTRONICA	46182	32,2%
	CIVIL	20356	14,2%
	ELECTRICA	6783	4,7%
	SISTEMAS	17598	12,3%

Tabla 6 Frecuencias y porcentajes de participación de las variables categóricas del estudio.

Como se observa en la tabla anterior, un perfil típico o más frecuente de un estudiante sería hombre, de estrato 3, que vive con sus padres, tiene como responsable a la madre, el nivel de educación es secundaria, vivienda propia y pagada, proviene de colegio privado de calendario A y se inscribe en el programa de Ingeniería Industrial. Para las variables continuas, se presenta la descripción estadística:

	Media	Mínimo	Máximo	Desviación típica	Moda	Varianza
Valor Matricula	4145059	2475000	6668970	872806	3694420	8,E11
Total ingreso	13103229	480000	60000000	21039732	7200000	4,E14
Edad al entrar	17	14	22	1	17	1
Icfes Matemáticas	48	26	92	7	51	50
Icfes Física	51	24	87	7	48	43
Icfes Lenguaje	56	26	90	7	56	46
AGEO -Análisis Geométrico	3,2	,0	5,0	,7	3,0	,5
PREC - Precálculo	3,0	,0	5,0	,7	3,0	,5
FFIS - Fund. Física	3,2	,0	5,0	,7	3,0	,5
ALLI- Algebra Lineal	3,1	,0	5,0	,7	3,0	,5
CALD - Cálculo Diferencial	3,0	,0	5,0	,7	3,0	,5
CALI - Cálculo Integral	3,1	,0	5,0	,7	3,1	,5
CALV - Cálculo Vectorial	3,0	,0	5,0	,7	3,0	,5
ECDI - Ecuaciones Diferenciales	3,1	,0	5,0	,7	3,0	,5
QUIM - Química	2,8	,0	4,6	,5	3,0	,3
BIOL - Biología	3,7	,0	4,7	,5	4,0	,3
FIS1 -Física 1	3,2	,0	5,0	,6	3,3	,3
FIS2 - Física 2	3,4	,0	5,0	,5	3,3	,3
FIS3 - Física 3	3,5	,0	4,8	,5	3,3	,2

Tabla 7 Estadísticos descriptivos de las variables continuas del estudio.

Estos valores continuos se normalizan a través de los valores tipificados del SPSS®, con el fin de manejar dimensiones similares con los demás datos y evitar que la varianza de los datos afecte los resultados.

Se observa que la edad mínima es 14 años y la máxima 22, que son relativamente menores que las edades que se venían presentando hace una década, muy seguramente afectados por las políticas de promoción automática que tenía el gobierno para el nivel de secundaria en todo el país (MEN, 2008).

Se hace un análisis de las correlaciones entre estas variables:

		Correlaciones					
		Edad al entrar	Valor Matrícula	Total ingreso	Icfes Matemáticas	Icfes Física	Icfes Lenguaje
Edad al entrar	Correlación de Pearson	1	,130**	,017**	,012**	,019**	,003
	Sig. (bilateral)		,000	,000	,000	,000	,189
	N	143401	143401	143401	143401	143401	143401
Valor Matrícula	Correlación de Pearson	,130**	1	,402**	,074**	,088**	,087**
	Sig. (bilateral)	,000		,000	,000	,000	,000
	N	143401	143401	143401	143401	143401	143401
Total ingreso	Correlación de Pearson	,017**	,402**	1	,091**	,089**	,051**
	Sig. (bilateral)	,000	,000		,000	,000	,000
	N	143401	143401	143401	143401	143401	143401
Icfes Matemáticas	Correlación de Pearson	,012**	,074**	,091**	1	,341**	,206**
	Sig. (bilateral)	,000	,000	,000		,000	,000
	N	143401	143401	143401	143401	143401	143401
Icfes Física	Correlación de Pearson	,019**	,088**	,089**	,341**	1	,165**
	Sig. (bilateral)	,000	,000	,000	,000		,000
	N	143401	143401	143401	143401	143401	143401
Icfes Lenguaje	Correlación de Pearson	,003	,087**	,051**	,206**	,165**	1
	Sig. (bilateral)	,189	,000	,000	,000	,000	
	N	143401	143401	143401	143401	143401	143401

** La correlación es significativa al nivel 0,01 (bilateral).

Tabla 8 Correlaciones de las variables continuas del estudio.

Hay bastantes relaciones significativas, lo cual muestra que es posible que sean variables significativas en el modelo, como efectivamente se comprueba más adelante. Se hace un análisis factorial exploratorio para ver qué tanto las variables pueden explicar la varianza de los datos, utilizando el método de máxima verosimilitud, puesto que la regresión logística multinomial usa también este método. Los resultados son:

KMO y prueba de Bartlett		
Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,542
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	62444,807
	gl	171
	Sig.	,000

Tabla 9 Resultados factor de significación del modelo.

Para la medida de KMO que contrasta si las correlaciones parciales entre las variables son pequeñas, da un valor de 0,542, lo cual indica que esta técnica de factor no es la mejor para el análisis de tales datos. No obstante, la prueba de esfericidad de Bartlett da significativa, porque su *p-value* es menor que 0,05; esta tiene como hipótesis nula que la matriz de correlaciones observada es una matriz identidad. El estadístico Bartlett se distribuye aproximadamente según el modelo de probabilidad Chi-cuadrado y es una transformación del determinante de la matriz de correlaciones (Hair et al.,1999). Como se rechaza la hipótesis, se concluye que el modelo podría ser adecuado para explicar los datos.

La varianza total explicada es:

Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	1,658	8,727	8,727	1,658	8,727	8,727
2	1,266	6,664	15,391	1,266	6,664	15,391
3	1,032	5,434	20,825	1,032	5,434	20,825
4	1,030	5,423	26,248	1,030	5,423	26,248
5	1,027	5,403	31,651	1,027	5,403	31,651
6	1,026	5,398	37,049	1,026	5,398	37,049
7	1,024	5,390	42,439	1,024	5,390	42,439
8	1,022	5,378	47,818	1,022	5,378	47,818
9	1,020	5,371	53,188	1,020	5,371	53,188
10	1,019	5,363	58,551	1,019	5,363	58,551
11	1,017	5,354	63,905	1,017	5,354	63,905
12	1,014	5,336	69,241	1,014	5,336	69,241
13	1,010	5,317	74,558	1,010	5,317	74,558
14	1,002	5,276	79,834	1,002	5,276	79,834
15	,977	5,143	84,977			
16	,863	4,544	89,521			
17	,758	3,988	93,509			
18	,656	3,454	96,963			
19	,577	3,037	100,000			

Método de extracción: Análisis de Componentes principales.

Tabla 10 Varianza explicada por los factores.

Como se observa, sólo se explica el 79,83% de la varianza de los datos; se espera que las otras variables contribuyan a explicar lo que hace falta del modelo. Sin embargo, indica que pueden ser variables significativas; en la siguiente tabla se presentan cómo se conformarían los factores.

Matriz de componentes*

	Componente													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Icfes Matemáticas	,610	-.462												
Icfes Física	,598	-.425												
Valor Matrícula	,592	,590												
Total ingreso	,570	,547												
Icfes Lenguaje	,469	-.338												
PREC - Precálculo			,669	-.634										
Edad al entrar			,350											
AGEO - Análisis Geométrico			,467	,783										
CALD - Cálculo Diferencial					-.728	-.507								
ALLI- Álgebra Lineal					,697	-.606								
FFIS - Fund. Física						,571	,734							
CALI - Cálculo Integral							-.390	-.813						
FIS1 - Física 1							-.356	,544	-.469	-.417				
CALV - Cálculo Vectorial									,868	-.315				
ECDI - Ecuaciones Diferenciales										,766	-.480			
FIS2 - Física 2											,864			
FIS3 - Física 3												,870	-.396	
QUIM - Química												,348	,891	
BIOL - Biología														,954

Método de extracción: Análisis de componentes principales.

a. 14 componentes extraídos

Tabla 11 Factores iniciales que explican el modelo.

Claramente, estas asignaturas son significativas y se pueden unir de a dos para formar un nuevo factor significativo, lo que da un total de catorce factores. Realmente no contribuye a la simplificación del estudio, pero es de esperarse porque son asignaturas claves en el desarrollo de la ingeniería. A continuación se presentan las pruebas de normalidad de los datos.

Prueba de Kolmogorov-Smirnov para una muestra

	N	Parámetros normales ^{a, b}		Diferencias más extremas			Z de Kolmogorov-Smirnov	Sig. asintót. (bilateral)
		Media	Desviación típica	Absoluta	Positiva	Negativa		
Edad al entrar	143401	16,67	,847	,228	,228	-.209	86,400	,000
Valor Matrícula	143401	4145128,36	872748,416	,106	,106	-.056	39,953	,000
Total ingreso	143401	13102619,62	21038033,90	,323	,323	-.276	122,196	,000
Icfes Matemáticas	143401	48,25	7,040	,090	,090	-.050	34,248	,000
Icfes Física	143401	50,67	6,549	,085	,085	-.063	32,131	,000
Icfes Lenguaje	143401	56,05	6,771	,074	,074	-.039	27,928	,000
Estrato	143401	3,08	,834	,290	,290	-.249	109,648	,000
Sexo	143401	1,66	,472	,425	,256	-.425	161,092	,000
Vive actualmente	143401	2,38	,812	,441	,441	-.318	167,125	,000
Calendario	143401	1,06	,294	,536	,536	-.420	202,966	,000
Naturaleza	143401	1,744	,4362	,465	,279	-.465	176,245	,000
Responsable	143401	1,16	,508	,523	,523	-.373	198,005	,000
Nivel educ. Resp	143401	2,57	1,124	,244	,244	-.191	92,241	,000
Tipo Vivienda	143401	1,83	1,187	,332	,332	-.241	125,676	,000
Estado	143401	1,81	,982	,312	,312	-.205	118,232	,000

a. La distribución de contraste es la Normal.
b. Se han calculado a partir de los datos.

Tabla 12 Prueba de Kolmogorov-Smirnov para la normalidad de los datos.

Como se observa en la tabla anterior, los datos dan significativos, es decir, el *p-value* es inferior a 0,05, con lo cual se rechaza la hipótesis nula de la prueba de Kolmogorov-Smirnov que señala que los datos de las variables tienen una distribución normal, por lo que se concluye que los datos de entrada no siguen una distribución normal.

Se presenta a continuación la prueba de homogeneidad de las varianzas, cuyo resultado es:

	Estadístico de Levene	gl1	gl2	Sig.
Edad al entrar	183,442	2	143398	,000
Valor Matrícula	26,538	2	143398	,000
Total ingreso	46,595	2	143398	,000
Icfes Matemáticas	65,424	2	143398	,000
Icfes Física	82,556	2	143398	,000
Icfes Lenguaje	10,715	2	143398	,000
Estrato	103,343	2	143398	,000
Sexo	1254,817	2	143398	,000
Vive actualmente	253,762	2	143398	,000
Calendario	628,683	2	143398	,000
Naturaleza	160,189	2	143398	,000
Responsable	1850,752	2	143398	,000
Nivel educ. Resp	126,734	2	143398	,000
Tipo Vivienda	275,046	2	143398	,000
Estado	1451,890	2	143398	,000

Tabla 13 Prueba de Levene. Homogeneidad de los datos.

La prueba de Levene tiene como hipótesis nula que las varianzas de las variables no son diferentes, es decir, que las variables son homogéneas (Hair et al., 1999); como se presenta en la tabla, todas dieron significativas y el valor del *p-value* fue inferior a 0,05, por lo que se rechaza la hipótesis nula. Con esto se concluye que las variables de los datos no son homogéneas.

Con base en los resultados anteriores y teniendo en cuenta que la variable dependiente de deserción es categórica, se define la regresión logística multinomial como la prueba adecuada para el conjunto de datos en estudio, cuyos resultados se presentan más adelante.

7.5. Análisis de valores perdidos

La regresión logística multinomial no permite que los datos tengan valores perdidos, por lo que se hace un análisis de estos datos para tratarlos.

Luego de analizar los datos con SPSS®, se obtienen los siguientes resultados:

	Perdidos		N válido	Media	Desviación típica
	N	Porcentaje			
Icfes Lenguaje	3032	2,1%	140369	56,06	6,771
Icfes Física	3032	2,1%	140369	50,66	6,547
Icfes Matemáticas	3032	2,1%	140369	48,24	7,017
Naturaleza	1171	,8%	142230		
Calendario	1171	,8%	142230		
Valor Matrícula	125	,1%	143276	4145120,29	872910,483
Edad al entrar	123	,1%	143278	16,67	,847

Tabla 14 Datos perdidos de la base de datos.

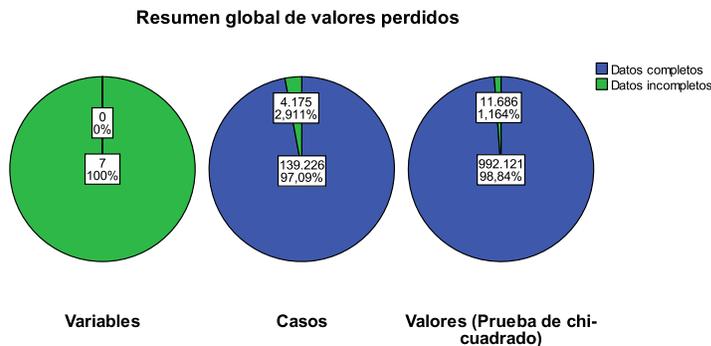
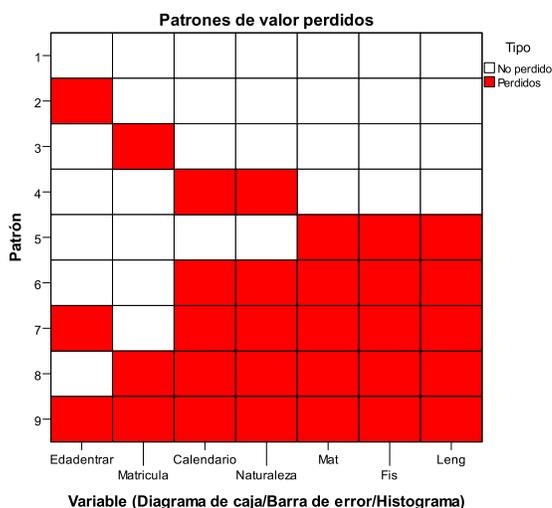


Figura 9 Resumen global de datos perdidos.

En la figura anterior se aprecian tres gráficas que presentan diferentes aspectos de los valores perdidos, como las variables, los casos y los valores.



En esta figura se muestran los patrones de los datos perdidos: el patrón 1 representa los casos de la base de datos que no tienen datos perdidos, el patrón 4 representa los casos que tienen datos perdidos en las variables Calendario (calendario del colegio, A, B, etc.) y Naturaleza (si el colegio es privado u oficial), y el patrón 9 representa los casos que tienen valores perdidos en Edad al entrar, Matrícula, Calendario, Naturaleza, Icfes de Matemáticas, Física y Lenguaje.

Este gráfico ayuda a identificar si el método de imputación para los datos debe ser monótonico o no y si los datos presentan un patrón monótonico. No todas las casillas perdidas y no perdidas del gráfico están contiguas, por lo que los datos no siguen un patrón monótono; en consecuencia, no se puede aplicar el método monótono para imputar los datos. Se aplica el método MCMC completamente condicional, un método iterativo basado

en las cadenas de Markov y Monte Carlo que se usa cuando los datos perdidos son arbitrarios, en tanto que el de regresión múltiple se utiliza para las variables continuas. Se corre el proceso con una sola imputación y se obtiene la base corregida para usarla con el proceso de regresión logística multinomial.

Especificaciones de imputación	
Método de imputación	Especificación condicional completa
Número de imputaciones	1
Modelo para variables de escala	Regresión lineal
Interacciones incluidas en modelos	Interacciones bidireccionales
Porcentaje máximo de valores perdidos	100,0%
Número máximo de parámetros del modelo de imputación.	100

Tabla 15 Especificaciones de la imputación.

	Papel en imputación		Valores imputados	
	Dependiente (Regresión logística)	Predictor	Mínimo	Máximo
Edad al entrar	Sí	Sí	15	23
Valor Matrícula	Sí	Sí	3500000	(ninguna)
Icfes Matemáticas	Sí	Sí	30	90
Icfes Física	Sí	Sí	30	90
Icfes Lenguaje	Sí	Sí	30	90
Calendario	Sí	Sí		
Naturaleza	Sí	Sí		

Tabla 16 Restricciones de las imputaciones.

Resultados de imputación	
Método de imputación	Especificación condicional completa
Iteraciones de método de especificación totalmente condicional	10
Variables dependientes	Imputado Edadentrar,Matricula,Mat,Fis,Leng, Calendario,Naturaleza
	No imputado (demasiados valores perdidos)
	No imputado (sin valores perdidos)
Secuencia de imputación	Edadentrar,Matricula,Mat,Fis,Leng, Calendario,Naturaleza

Tabla 17 Resultados de la imputación.

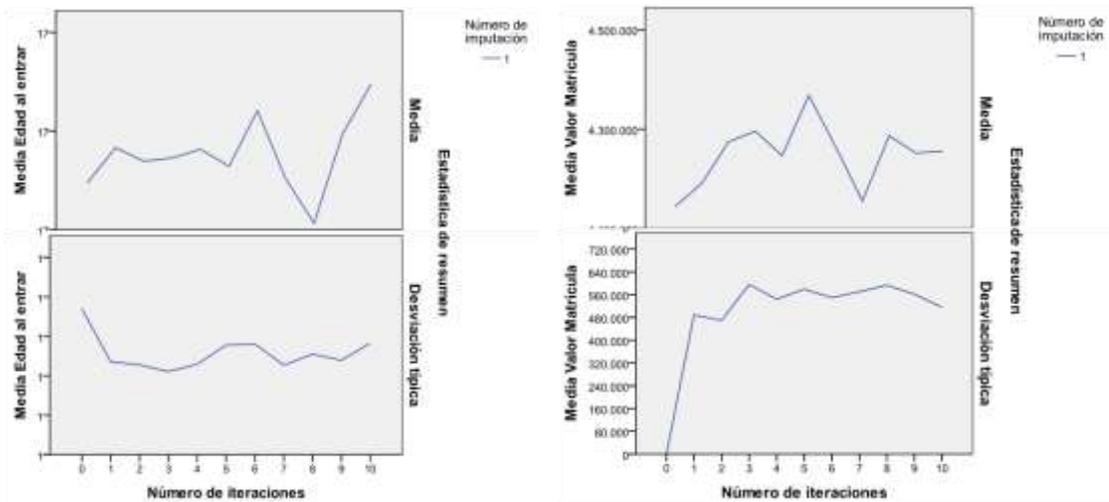


Figura 10 Edad al entrar y valor de matrícula imputados.

Según las figuras, los valores imputados para las variables de Edad al entrar, Valor de la Matrícula, Icfes de Matemáticas, Física y Lenguaje no tienen un patrón sobre el comportamiento de estas.

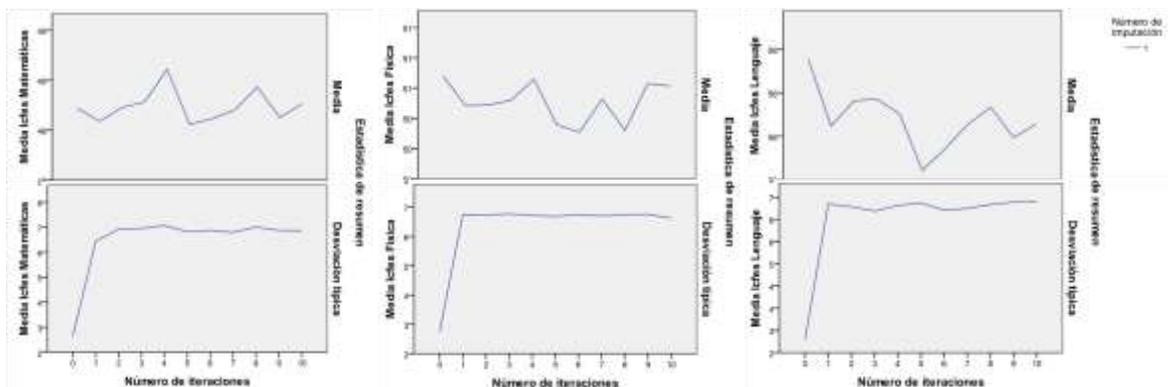


Figura 11 Icfes de Matemáticas, Física y Lenguaje imputados.

Se han modificado algunas escalas para hacer visible las líneas y poder observar su comportamiento.

7.6. Prueba de independencia Chi-cuadrado

La probabilidad de que un estudiante cambie de estado, esto es, de que apruebe el semestre, permanezca o se retire, es independiente de sus semestres anteriores.

Pruebas de chi-cuadrado de Pearson

		Sexo			
AGEO	Chi cuadrado	54,265	QUIM	Chi cuadrado	55,428
	gl	49		gl	49
	Sig.	,281		Sig.	,346
PREC	Chi cuadrado	46,219	ECDI	Chi cuadrado	58,662
	gl	49		gl	47
	Sig.	,587		Sig.	,118
FFIS	Chi cuadrado	51,375	BIOL	Chi cuadrado	53,602
	gl	48		gl	28
	Sig.	,343		Sig.	,207
CALD	Chi cuadrado	56,700	AGEO	Chi cuadrado	54,265
	gl	49		gl	49
	Sig.	,210		Sig.	,281
ALLI	Chi cuadrado	69,145	FIS2	Chi cuadrado	47,825
	gl	49		gl	40
	Sig.	,153		Sig.	,185
CALI	Chi cuadrado	59,876	FIS3	Chi cuadrado	41,924
	gl	49		gl	34
	Sig.	,137		Sig.	,165
CALV	Chi cuadrado	41,587	Los resultados se basan en filas y columnas no vacías de cada subtabla más al interior.		
	gl	48			
	Sig.	,731			

7.7. Programa para calcular la variable dependiente

Existe un problema conceptual y es que un estudiante puede cursar periodos de tiempo y aun así no avanzar en los semestres académicos; en otras palabras, una persona puede estudiar tres años en la universidad y estar en segundo semestre. Esto puede pasar por varias razones; por ejemplo, el estudiante sólo paga media matrícula y no ve sino la mitad de la carga, o el estudiante pierde materias o las cancela, y entonces no avanza. Este punto, clave para el estudio, marca una diferencia frente al tradicional conteo que se hace de que sólo se mira si el estudiante pasa o no el periodo de tiempo, o si se demora 15 o 20 semestres para graduarse, pero este manejo no es adecuado, porque realmente son diez semestres académicos, no hay más; lo que sucede es que el tiempo que se demora en cursar estos diez semestres no está sincronizado. En este estudio es diferente un semestre académico de un periodo de tiempo que el estudiante curse.

Para cubrir este aspecto ya explicado, se creó un programa que identifica si la persona se retira. En este caso, el programa le asigna el valor de 1 a la variable dependiente sólo en los registros del último periodo que esté cursando; esto en razón de que puede que se retire,

aunque en el transcurso de sus estudios aprobó varios semestres antes o perdió alguno, pero no se fue sino varios periodos después. Igualmente, con base en las notas y el número de créditos de una materia, el programa indica si aprobó o no el periodo. Para esto calcula el promedio ponderado de las asignaturas que cursó el estudiante, y si este obtiene un promedio inferior a 3,0 significa que pierde el semestre académico y a la variable dependiente se le asigna el valor de 2; si el promedio obtenido es mayor o igual que 3,0, quiere decir que aprobó el semestre y se le asigna un valor de 3 a la variable dependiente.

Luego el *software* suma los créditos aprobados y calcula el semestre en el que el estudiante se encuentra, teniendo en cuenta el programa de ingeniería al que pertenece, se calcula con una operación de proporción (regla de tres) y se redondea. Cabe destacar que los semestres siempre son consecutivos y van de menor a mayor, no es posible estar en un semestre dado y al siguiente periodo estar en un semestre menor, si el estudiante pierde no se aumenta el semestre quedando como semestre el valor anterior, pero el valor del semestre nunca se disminuye.

Existen condiciones especiales que se tuvieron en cuenta, tales como casos de estudiantes que han perdido dos, tres o cuatro veces un semestre académico; periodos intersemestrales que se integran a los semestres porque máximo se pueden ver ocho créditos, que equivalen a dos materias; diferentes números de créditos totales que se deben cursar según el programa de Ingeniería escogido, y estudiantes que hacen doble programa.

7.8. Regresión logística multinomial

Una vez que se tiene la base depurada, se separa por semestres académicos para correr el proceso de regresión logística multinomial. Cada semestre se divide en dos muestras, una para entrenamiento del modelo y otra para hacer la prueba de este, y como ya se definió que se va a usar la regresión logística multinomial, no se requiere hacer una base para la validación.

Los porcentajes empleados fueron 60% para entrenamiento y 40% para la prueba; sin embargo, cuando los porcentajes de la variable dependiente en los valores de deserción

fueron disminuyendo se ajustó la base de entrenamiento, con el fin de obtener un modelo que encontrara estos casos. Luego se procede a correr el modelo y se analiza qué variables son significativas, con el objeto de dejarlas dentro del modelo final. Se hacen varias corridas, hasta que se obtiene un modelo óptimo. Después que está listo, se corre en la muestra de datos de prueba para ver qué tan correcto sale el modelo; igualmente, se hace un análisis de su validez interna para asegurar el óptimo válido (Hair et al., 1999).

Una vez obtenidos los betas para cada semestre, se realiza un programa en Visual Basic de Visual Studio 2008®, que opera los betas con una matriz de perfil de estudiante que da valores a las variables presentadas en la tabla anterior y genera la matriz de transición para continuar con el análisis. El trabajo desarrollado semestre a semestre, se encuentra en detalle en el anexo 1.

7.9. Matrices de transición de Markov

Para obtener las matrices de transición de Markov se escogen unos perfiles con el propósito de conocer su comportamiento y el cambio que tienen cuando se varía alguna de las variables; las demás se dejan constantes, para ver los efectos en el valor de la variable de análisis.

Se construye un perfil con los valores más frecuentes de los datos y se modifica una variable que se quiere analizar para ver cómo es su efecto en las probabilidades de deserción y de graduación. La primera variable que se debe cambiar es el nivel de educación del responsable.

Nivel Edu	Estrato	Sexo	Vive actualmente	Responsable	Tipo de vivienda	Edad	Ingreso
Primaria	3	Hombre	Con padres	Madre	Propia y pagada	18	Medio

El programa arroja la siguiente matriz de probabilidades, que se puede exportar a Access, y luego a Excel, para continuar con las operaciones.

Tabla 18 Matriz de probabilidades de transición, perfil 1.

Se presenta en Excel, para facilitar su visibilidad.

Id	sem1	sem2	sem3	sem4	sem5	sem6	sem7	sem8	sem9	sem10	Graduan	Retiran	Suma
Sem1	0,169260405	0,7054511	0	0	0	0	0	0	0	0	0	0,1252885	1
Sem2	0	0,16986367	0,72905502	0	0	0	0	0	0	0	0	0,10108131	1
Sem3	0	0	0,16154708	0,71063085	0	0	0	0	0	0	0	0,12782207	1
Sem4	0	0	0	0,14534838	0,73946907	0	0	0	0	0	0	0,11518255	1
Sem5	0	0	0	0	0,14153955	0,77051237	0	0	0	0	0	0,08794808	1
Sem6	0	0	0	0	0	0,12471182	0,80476729	0	0	0	0	0,07052089	1
Sem7	0	0	0	0	0	0	0,0955315	0,85124625	0	0	0	0,05322225	1
Sem8	0	0	0	0	0	0	0	0,12551421	0,79352097	0	0	0,08096482	1
Sem9	0	0	0	0	0	0	0	0	0,0306838	0,95486091	0	0,01445529	1
Sem10	0	0	0	0	0	0	0	0	0	0,02668472	0,95990416	0,01341111	1
Graduan	0	0	0	0	0	0	0	0	0	0	1	0	1
Retiran	0	0	0	0	0	0	0	0	0	0	0	1	1

Tabla 19 Probabilidades de transición, perfil 1, en Excel.

Se opera esta matriz de transición para obtener la matriz fundamental, $M = (I - Q)^{-1}$, que nos da el tiempo de permanencia de una persona desde que ingresa a la universidad hasta que se gradúa. Para este perfil da una duración de 6,82610 años.

$M = (I - Q)^{-1}$	1	2	3	4	5	6	7	8	9	10	Tiempo de permanencia
1	1,20375	1,02295	0,88948	0,73959	0,63707	0,56081	0,49899	0,48573	0,39764	0,3901	6,82610
2	0	1,20462	1,04745	0,87094	0,75022	0,66041	0,58761	0,572	0,468261	0,45938	
3	0	0	1,19267	0,99169	0,85423	0,75198	0,66909	0,6513	0,533184	0,52307	
4	0	0	0	1,17007	1,00788	0,88724	0,78943	0,76846	0,629088	0,61716	
5	0	0	0	0	1,16488	1,02544	0,9124	0,88815	0,727077	0,71329	
6	0	0	0	0	0	1,14248	1,01654	0,98953	0,810067	0,79471	
7	0	0	0	0	0	0	1,10562	1,07624	0,881053	0,86435	
8	0	0	0	0	0	0	0	1,14353	0,936139	0,91839	
9	0	0	0	0	0	0	0	0	1,031655	1,01209	
10	0	0	0	0	0	0	0	0	0	1,02742	

Tabla 20 Matriz M para el primer perfil analizado.

Luego se opera $M \cdot R$. Donde la matriz R es la parte sombreada de la tabla 20.

Los resultados son:

M*R = Nivel de educación Primaria		
	Gradúan	Retiran
1	0,37446	0,62554
2	0,44096	0,55904
3	0,50210	0,49790
4	0,59241	0,40759
5	0,68469	0,31531
6	0,76284	0,23716
7	0,82969	0,17031
8	0,88157	0,11843
9	0,97151	0,02849
10	0,98622	0,01378

Así, la probabilidad de que se gradúe entrando en primer semestre es de 0,37446 y la probabilidad de que se retire es de 0,62554.

Para el siguiente nivel de educación, que es secundaria, los resultados son:

Nivel Edu	Estrato	Sexo	Vive actualmente	Responsable	Tipo de vivienda	Edad	Ingreso
Secundaria	3	Hombre	Con padres	Madre	Propia y pagada	18	Medio

Id	sem1	sem2	sem3	sem4	sem5	sem6	sem7	sem8	sem9	sem10	Graduan	Retiran	Suma
Sem1	0,168181732	0,70772888	0	0	0	0	0	0	0	0	0	0,12408939	1
Sem2	0	0,16591286	0,73633267	0	0	0	0	0	0	0	0	0,09775447	1
Sem3	0	0	0,13313506	0,76959439	0	0	0	0	0	0	0	0,09727055	1
Sem4	0	0	0	0,12795675	0,77533328	0	0	0	0	0	0	0,09670998	1
Sem5	0	0	0	0	0,14315978	0,76754102	0	0	0	0	0	0,08929921	1
Sem6	0	0	0	0	0	0,16874938	0,72538522	0	0	0	0	0,10586539	1
Sem7	0	0	0	0	0	0	0,10456593	0,83612504	0	0	0	0,05930903	1
Sem8	0	0	0	0	0	0	0	0,12536195	0,79379987	0	0	0,08083818	1
Sem9	0	0	0	0	0	0	0	0	0,07869298	0,88113231	0	0,04017471	1
Sem10	0	0	0	0	0	0	0	0	0	0,09157895	0,85686075	0,0515603	1
Graduan	0	0	0	0	0	0	0	0	0	0	0	1	0
Retiran	0	0	0	0	0	0	0	0	0	0	0	0	1

Tabla 21 Matriz de probabilidades de transición, perfil 2.

Se opera esta matriz de transición para obtener la matriz fundamental, $M = (I - Q)^{-1}$, que nos da el tiempo de permanencia de una persona desde que ingresa hasta que se gradúa.

$M=(I-Q)^{-1}$	1	2	3	4	5	6	7	8	9	10	Tiempo de permanencia
1	1,20219	1,02006	0,86646	0,76467	0,69193	0,6389	0,51757	0,49478	0,4263	0,41349	7,03635
2	0	1,19892	1,01838	0,89874	0,813249	0,75092	0,60831	0,58153	0,50105	0,48599	
3	0	0	1,15358	1,01806	0,921215	0,85061	0,68907	0,65873	0,56756	0,55051	
4	0	0	0	1,14673	1,037649	0,95812	0,77617	0,74199	0,6393	0,6201	
5	0	0	0	0	1,167079	1,07763	0,87298	0,83454	0,71904	0,69744	
6	0	0	0	0	0	1,20301	0,97455	0,93164	0,8027	0,77859	
7	0	0	0	0	0	0	1,11678	1,0676	0,91985	0,89222	
8	0	0	0	0	0	0	0	1,14333	0,9851	0,9555	
9	0	0	0	0	0	0	0	0	1,08541	1,05281	
10	0	0	0	0	0	0	0	0	0	1,10081	

Da un tiempo de permanencia de 7,03635.

M*R = Nivel de educación Secundaria		
	Gradúan	Retiran
1	0,35431	0,64569
2	0,41643	0,58357
3	0,47171	0,52829
4	0,53134	0,46866
5	0,59761	0,40239
6	0,66714	0,33286
7	0,76450	0,23550
8	0,81873	0,18127
9	0,90211	0,09789
10	0,94324	0,05676

Da una probabilidad de graduarse de 0,35431 y de retirarse de 0,64569.

El detalle de los resultados para cada perfil se puede observar en el anexo 2.

A continuación se presenta un cuadro resumen de los resultados obtenidos.

Nivel Edu	Estrato	Sexo	Vive actualmente	Responsable	Tipo de vivienda	Edad	Ingreso
Cambia	3	Hombre	Con padres	Madre	Propia y pagada	18	Medio

Nivel Edu	Tiempo de permanencia	Probabilidad de Graduarse	Probabilidad de Retirarse
Primaria	6,82610	0,37446	0,62554
Secundaria	7,03635	0,35431	0,64569
Técnica	6,91486	0,37657	0,62343
Universitaria o más	7,09201	0,37165	0,62835

Se analiza este mismo cambio de variable de nivel de educación para las mujeres.

Nivel Edu	Estrato	Sexo	Vive actualmente	Responsable	Tipo de vivienda	Edad	Ingreso
Cambia	3	Mujer	Con padres	Madre	Propia y pagada	18	Medio

Nivel Edu	Tiempo de permanencia	Probabilidad de Graduarse	Probabilidad de Retirarse
Primaria	7,38154	0,45625	0,54375
Secundaria	7,47627	0,42694	0,57306
Técnica	7,54102	0,47485	0,52515
Universitaria o más	7,51190	0,45790	0,54210

Según los resultados, las mujeres tienen mayor probabilidad de graduarse; no obstante, el tiempo de permanencia es mayor que el de los hombres. Se corre la prueba de Tamhane para analizar las diferencias de medias de la probabilidad de graduarse con los diferentes niveles de educación del responsable del estudiante.

Comparaciones múltiples

Probabilidad de Graduarse
Tamhane

(I) Nivel de educación	(J) Nivel de educación	Diferencia de medias (I-J)	Error típico	Sig.	Intervalo de confianza al 95%	
					Límite inferior	Límite superior
Primaria	Secundaria	-,015248727	,077880941	1,000	-,23141112	,20091366
	Técnico	-,012411907	,077689357	1,000	-,22804926	,20322545
	Universidad o más	-,019813473	,077666542	1,000	-,23538835	,19576140
Secundaria	Primaria	,015248727	,077880941	1,000	-,20091366	,23141112
	Técnico	,002836820	,075617771	1,000	-,20701296	,21268660
	Universidad o más	-,004564746	,075594330	1,000	-,21434956	,20522006
Técnico	Primaria	,012411907	,077689357	1,000	-,20322545	,22804926
	Secundaria	-,002836820	,075617771	1,000	-,21268660	,20701296
	Universidad o más	-,007401566	,075396936	1,000	-,21663820	,20183506
Universidad o más	Primaria	,019813473	,077666542	1,000	-,19576140	,23538835
	Secundaria	,004564746	,075594330	1,000	-,20522006	,21434956
	Técnico	,007401566	,075396936	1,000	-,20183506	,21663820

Tabla 22 Prueba Tamhane para el nivel de educación del responsable del estudiante.

Como se observa en la tabla anterior, para todos los niveles de educación del responsable del estudiante el valor de significancia dio superior a 0,05, por lo que se concluye que todos los promedios comparados de los diferentes niveles de educación del responsable del estudiante no difieren significativamente.

El segundo perfil seleccionado es cambiar el estrato y se dejan todas las demás variables constantes:

Estrato	Nivel Edu	Sexo	Vive actualmente	Responsable	Tipo de vivienda	Edad	Ingreso
Cambia	Secundaria	Hombre	Con padres	Madre	Propia y pagada	18	Medio

Estrato	Tiempo de permanencia	Probabilidad de Graduarse	Probabilidad de Retirarse
1	7,09440	0,43647	0,56353
2	7,35458	0,36441	0,63559
3	7,28773	0,37173	0,62827
4	7,57067	0,43524	0,56476
5	6,89366	0,29290	0,70710
6	7,53435	0,41421	0,58579

El estrato 1 se demora mucho menos tiempo en graduarse en comparación con los otros estratos, y su probabilidad de graduarse es mayor que la de los demás estratos socio económicos. Esto puede deberse al compromiso que tienen estos estudiantes con la oportunidad que se les presenta de poder cursar una ingeniería en una universidad privada de prestigio, pues en esta sociedad el nivel de educación se convierte en un ascensor social (Donoso & Schiefelbein, 2007).

Estrato				
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1	2220	1,5	1,5
	2	28576	19,9	21,5
	3	77200	53,8	75,3
	4	27911	19,5	94,8
	5	6432	4,5	99,3
	6	1062	,7	100,0
Total		143401	100,0	

Tabla 23 Frecuencias de la variable estrato socioeconómico.

Cuando se revisa la proporción de participación, se observa que apenas el 1,5% de la población es de estrato 1 y que la participación del estrato 5 en la Escuela es mínima, por lo que los datos obtenidos no son concluyentes. Los estudiantes de escasos recursos tienen más empeño e intentan graduarse más rápidamente, por los costos asociados a estar en una universidad privada (Choy et al., 2000), mientras que los estudiantes de estrato 5 no tienen el mismo afán, ya que para ellos el dinero no es una necesidad.

Se corre la prueba de Tamhane para analizar las diferencias de medias de la probabilidad de graduarse con los diferentes estratos socioeconómicos del estudiante.

Comparaciones múltiples						
Probabilidad de graduarse Tamhane						
(I) Estrato socioeconómico	(J) Estrato socioeconómico	Diferencia de medias (I-J)	Error típico	Sig.	Intervalo de confianza al 95%	
					Límite inferior	Límite superior
1	2	-.023667165	.075372430	1,000	-.25910092	.21176659
	3	.010330754	.074975119	1,000	-.22385312	.24451463
	4	-.050859330	.069157907	1,000	-.26718396	.16546530
	5	.028839735	.085231387	1,000	-.23842278	.29610225
	6	-.041569188	.073991012	1,000	-.27267165	.18953327
2	1	.023667165	.075372430	1,000	-.21176659	.25910092
	3	.033997919	.076207524	1,000	-.20402902	.27202486
	4	-.027192165	.070492088	1,000	-.24784430	.19345997
	5	.052506901	.086317482	1,000	-.21790592	.32291972
	6	-.017902022	.075239539	1,000	-.25292463	.21712058
3	1	-.010330754	.074975119	1,000	-.24451463	.22385312
	2	-.033997919	.076207524	1,000	-.27202486	.20402902
	4	-.061190084	.070067109	.999	-.28046087	.15808071
	5	.018508982	.085970768	1,000	-.25089403	.28791200
	6	-.051899941	.074841522	1,000	-.28566929	.18186941
4	1	.050859330	.069157907	1,000	-.16546530	.26718396
	2	.027192165	.070492088	1,000	-.19345997	.24784430
	3	.061190084	.070067109	.999	-.15808071	.28046087
	5	.079699066	.080947641	.998	-.17555873	.33495686
	6	.009290143	.069013051	1,000	-.20656629	.22514658
5	1	-.028839735	.085231387	1,000	-.29610225	.23842278
	2	-.052506901	.086317482	1,000	-.32291972	.21790592
	3	-.018508982	.085970768	1,000	-.28791200	.25089403
	4	-.079699066	.080947641	.998	-.33495686	.17555873
	6	-.070408923	.085113891	1,000	-.33733296	.19651512
6	1	.041569188	.073991012	1,000	-.18953327	.27267165
	2	.017902022	.075239539	1,000	-.21712058	.25292463
	3	.051899941	.074841522	1,000	-.18186941	.28566929
	4	-.009290143	.069013051	1,000	-.22514658	.20656629
	5	.070408923	.085113891	1,000	-.19651512	.33733296

Tabla 24 Prueba Tamhane para el estrato socioeconómico.

Como se observa en la tabla anterior, para todos estratos socio económicos del estudiante el valor de significancia dio superior a 0,05, por lo que se concluye que los promedios comparados de la probabilidad de graduarse no difieren significativamente debido al los estratos del estudiante.

La siguiente variable de cambio es edad, que varía de 15 a 22 años.

Edad	Nivel Edu	Estrato	Sexo	Vive actualmente	Responsable	Tipo de vivienda	Ingreso
Cambia	Secundaria	3	Hombre	Con padres	Madre	Propia y pagada	Medio

Edad	Tiempo de permanencia	Probabilidad de Graduarse	Probabilidad de Retirarse
15	7,33964	0,36683	0,63317
16	7,28773	0,37173	0,62827
17	7,21593	0,36765	0,63235
18	7,03635	0,35431	0,64569
19	6,32937	0,33318	0,66682
20	6,64383	0,32261	0,67739
21	6,32937	0,30335	0,69665
22	6,19514	0,29624	0,70376

Aunque la probabilidad de deserción es menor en los estudiantes más jóvenes, su tiempo de permanencia en la universidad es mucho más alto; un estudiante de 22 años se demora en promedio casi un año menos que un estudiante de 18 años. En Alemania, por ejemplo, debido al sistema de educación que se tiene, los estudiantes tienen mínimo 18 años al iniciar la educación superior (Lamb et al., 2011).

Un estudiante de mayor edad tiene menor probabilidad de graduarse, pero el que se queda hace su carrera universitaria en menos tiempo; esto puede deberse a la madurez y a que es posible que se haya concientizado de que a la menor falla puede perder su oportunidad de terminar la carrera. Se corre la prueba de Tamhane para analizar las diferencias de medias de la probabilidad de graduarse con las diferentes edades al entrar a la carrera de los estudiantes.

Comparaciones múltiples

Probabilidad de graduarse
Tamhane

(I) Edad al entrar	(J) Edad al entrar	Diferencia de medias (I-J)	Error típico	Sig.	Intervalo de confianza al 95%								
					Límite inferior	Límite superior							
15	16	.029367728	.107106841	1,000	-.35794519	.41668065	19	15	-.081327135	.112011431	1,000	-.46876328	.32610901
	17	.036548025	.107601506	1,000	-.35276612	.42586217		16	-.051959407	.117950994	1,000	-.48295213	.37903332
	18	.048939948	.109193312	1,000	-.34685934	.44473923		17	-.044779111	.118400362	1,000	-.47732720	.38776898
	19	.081327135	.112011431	1,000	-.32610901	.48876328		18	-.032387187	.119848825	1,000	-.47001491	.40524054
	20	.093758008	.112967456	1,000	-.31766701	.50518302		20	.012430872	.123297233	1,000	-.43766172	.46252346
	21	.125204124	.115777702	1,000	-.29805400	.54846225		21	.043876988	.125877074	1,000	-.41584362	.50359760
	22	.138627533	.116700234	1,000	-.28854750	.56580257		22	.057300398	.126726107	1,000	-.40563951	.52024031
16	15	-.029367728	.107106841	1,000	-.41668065	.35794519	20	15	-.093758008	.112967456	1,000	-.50518302	.31766701
	17	.007180297	.113771528	1,000	-.40812977	.42249036		16	-.064390280	.118859255	1,000	-.49887427	.37009371
	18	.019572221	.115278167	1,000	-.40131425	.44045869		17	-.057209983	.119305202	1,000	-.49322023	.37880026
	19	.051959407	.117950994	1,000	-.37903332	.48295213		18	-.044818059	.120742811	1,000	-.48581543	.39617931
	20	.064390280	.118859255	1,000	-.37009371	.49887427		19	-.012430872	.123297233	1,000	-.46252346	.43766172
	21	.095836396	.121533339	1,000	-.34907797	.54075076		21	.031446116	.126726107	1,000	-.43128105	.49417329
	22	.109259805	.122412503	1,000	-.33912971	.55764932		22	.044869526	.127571908	1,000	-.42102861	.51076767
17	15	-.036548025	.107601506	1,000	-.42586217	.35276612	21	15	-.125204124	.115777702	1,000	-.54846225	.29805400
	16	-.007180297	.113771528	1,000	-.42249036	.40812977		16	-.095836396	.121533339	1,000	-.54075076	.34907797
	18	.012391924	.115737913	1,000	-.41013843	.43492228		17	-.088656099	.121969510	1,000	-.53501377	.35770157
	19	.044779111	.118400362	1,000	-.38776898	.47732720		18	-.076264175	.123376078	1,000	-.52734508	.37481673
	20	.057209983	.119305202	1,000	-.37880026	.49322023		19	-.043876988	.125877074	1,000	-.50359760	.41584362
	21	.088656099	.121969510	1,000	-.35770157	.53501377		20	-.031446116	.126726107	1,000	-.49417329	.43128105
	22	.102079509	.122845552	1,000	-.34772693	.55186595		22	.013423410	.130066989	1,000	-.46137882	.48822564
18	15	-.048939948	.109193312	1,000	-.44473923	.34685934	22	15	-.138627533	.116700234	1,000	-.56580257	.28854750
	16	-.019572221	.115278167	1,000	-.44045869	.40131425		16	-.109259805	.122412503	1,000	-.55764932	.33912971
	17	-.012391924	.115737913	1,000	-.43492228	.41013843		17	-.102079509	.122845552	1,000	-.55186595	.34772693
	19	.032387187	.119848825	1,000	-.40524054	.47001491		18	-.089687585	.124242202	1,000	-.54413322	.36475805
	20	.044818059	.120742811	1,000	-.39617931	.48581543		19	-.057300398	.126726107	1,000	-.52024031	.40563951
	21	.076264175	.123376078	1,000	-.37481673	.52734508		20	-.044869526	.127571908	1,000	-.51076767	.42102861
	22	.089687585	.124242202	1,000	-.36475805	.54113322		21	-.013423410	.130066989	1,000	-.48822564	.46137882

Tabla No. 25 Prueba Tamhane para la edad al entrar.

Como se observa en la tabla anterior, para todos las edades del estudiante el valor de significancia dio superior a 0,05, por lo que se concluye que la edad al entrar del estudiante no es estadísticamente significativo en la probabilidad de graduarse.

La siguiente variable por cambiar fue con quién vive actualmente.

Vive actualmente	Edad	Nivel Edu	Estrato	Sexo	Responsable	Tipo de vivienda	Ingreso
Cambia	18	Secundaria	3	Hombre	Madre	Propia y pagada	Medio

Vive actualmente	Tiempo de permanencia	Probabilidad de Graduarse	Probabilidad de Retirarse
Vive solo	5,70772	0,09758	0,90242
Vive con padres	7,21593	0,36765	0,63235
Vive con madre	7,31192	0,38149	0,61851
Vive con padre	6,95181	0,33209	0,66791
Otro	7,34126	0,37329	0,62671

La probabilidad de graduarse cuando el estudiante vive solo es apenas de 0,09758, lejos de las probabilidades de graduarse en los demás casos. El estar solo puede implicar no tener un apoyo emocional que lo respalde en las decisiones cotidianas (A. W. Astin, 1977; Tinto, 2010; Spady, 1971).

Se corre la prueba de Tamhane para analizar las diferencias de medias de la probabilidad de graduarse cambiando con vive actualmente el estudiante.

Comparaciones múltiples

Probabilidad de graduarse
Tamhane

(I) Con quién vive	(J) Con quién vive	Diferencia de medias (I-J)	Error típico	Sig.	Intervalo de confianza al 95%	
					Límite inferior	Límite superior
Vive solo	Vive con padres	-,150820715	,152562014	,984	-,64995059	,34830916
	Vive con Madre	-,163353667	,150439097	,970	-,65811492	,33140759
	Vive con Padre	-,124408700	,155515164	,997	-,63006955	,38125215
	Otro	-,163353667	,150439097	,970	-,65811492	,33140759
Vive con padres	Vive solo	,150820715	,152562014	,984	-,34830916	,64995059
	Vive con Madre	-,012532952	,111386369	1,000	-,36755298	,34248708
	Vive con Padre	,026412015	,118152307	1,000	-,35025287	,40307690
	Otro	-,012532952	,111386369	1,000	-,36755298	,34248708
Vive con Madre	Vive solo	,163353667	,150439097	,970	-,33140759	,65811492
	Vive con padres	,012532952	,111386369	1,000	-,34248708	,36755298
	Vive con Padre	,038944967	,115398100	1,000	-,32940820	,40729814
	Otro	,000000000	,108460486	1,000	-,34556807	,34556807
Vive con Padre	Vive solo	,124408700	,155515164	,997	-,38125215	,63006955
	Vive con padres	-,026412015	,118152307	1,000	-,40307690	,35025287
	Vive con Madre	-,038944967	,115398100	1,000	-,40729814	,32940820
	Otro	-,038944967	,115398100	1,000	-,40729814	,32940820
Otro	Vive solo	,163353667	,150439097	,970	-,33140759	,65811492
	Vive con padres	,012532952	,111386369	1,000	-,34248708	,36755298
	Vive con Madre	,000000000	,108460486	1,000	-,34556807	,34556807
	Vive con Padre	,038944967	,115398100	1,000	-,32940820	,40729814

Tabla No. 26 Prueba Tamhane para con quién vive el estudiante.

Como se observa en la tabla anterior, para todos los cambios de con quién vive el estudiante el valor de significancia dio superior a 0,05, por lo que se concluye que esta variable no es estadísticamente significativa en la probabilidad de graduarse.

La siguiente variable de cambio fue el tipo de vivienda.

Tipo de vivienda	Vive actualmente	Edad	Nivel Edu	Estrato	Sexo	Responsable	Ingreso
Cambia	Con padres	18	Secundaria	3	Hombre	Madre	Medio

Tipo de vivienda	Tiempo de permanencia	Probabilidad de Graduarse	Probabilidad de Retirarse
Propia y pagada	7,18093	0,36352	0,63648
Propia y deuda	6,55606	0,20643	0,79357
Arrendada	6,27025	0,16226	0,83774
Familiar	7,26611	0,31540	0,68460
Otra	6,84419	0,28492	0,71508

Esto demuestra la importancia en la seguridad que los estudiantes perciben de su entorno social. Se corre la prueba de Tamhane para analizar las diferencias de medias de la probabilidad de graduarse cambiando el tipo de vivienda del estudiante.

Comparaciones múltiples

Probabilidad de graduarse
Tamhane

(I) Tipo de vivienda	(J) Tipo de vivienda	Diferencia de medias (I-J)	Error típico	Sig.	Intervalo de confianza al 95%	
					Límite inferior	Límite superior
Propia y pagada	Propia y deuda	,028812414	,136073169	1,000	-,40529134	,46291617
	Arendada	,057394157	,141471874	1,000	-,39502309	,50981140
	Familiar	-,042192275	,123885575	1,000	-,43733282	,35294827
	Otro	,000000000	,129285887	1,000	-,41192029	,41192029
Propia y deuda	Propia y pagada	-,028812414	,136073169	1,000	-,46291617	,40529134
	Arendada	,028581743	,147700229	1,000	-,44231087	,49947435
	Familiar	-,071004688	,130953054	1,000	-,49018097	,34817160
	Otro	-,028812414	,136073169	1,000	-,46291617	,40529134
Arendada	Propia y pagada	-,057394157	,141471874	1,000	-,50981140	,39502309
	Propia y deuda	-,028581743	,147700229	1,000	-,49947435	,44231087
	Familiar	-,099586431	,136554334	,998	-,53839732	,33922446
	Otro	-,057394157	,141471874	1,000	-,50981140	,39502309
Familiar	Propia y pagada	,042192275	,123885575	1,000	-,35294827	,43733282
	Propia y deuda	,071004688	,130953054	1,000	-,34817160	,49018097
	Arendada	,099586431	,136554334	,998	-,33922446	,53839732
	Otro	,042192275	,123885575	1,000	-,35294827	,43733282
Otro	Propia y pagada	,000000000	,129285887	1,000	-,41192029	,41192029
	Propia y deuda	,028812414	,136073169	1,000	-,40529134	,46291617
	Arendada	,057394157	,141471874	1,000	-,39502309	,50981140
	Familiar	-,042192275	,123885575	1,000	-,43733282	,35294827

Tabla No. 27 Prueba Tamhane para el tipo de vivienda del estudiante.

Se puede observar en la tabla anterior, para todos los cambios de con quién vive el estudiante el valor de significancia dio superior a 0,05, por lo que se concluye que esta variable no es estadísticamente significativa en la probabilidad de graduarse

La última variable que se cambia es la nota de las asignaturas de Ciencias Básicas. Se prueban dos perfiles: a uno se le pone a repetir algunas materias de matemáticas en los primeros semestres y al otro se le ponen notas aprobatorias en sus materias de los primeros semestres. Los resultados son:

Notas	Tiempo de permanencia	Probabilidad de Graduarse	Probabilidad de Retirarse
Notas bajas-repite	7,44359	0,20221	0,79779
Notas altas-no repite	6,29757	0,36225	0,63775

Como se puede observar, la probabilidad de graduarse es mucho mayor en el perfil que obtiene notas altas y que no pierde, frente al perfil que pierde asignaturas de Ciencias Básicas; igualmente, el tiempo de permanencia es menor en comparación con el otro perfil. Es una diferencia bastante alta, de casi 0,16 puntos, lo que lleva a pensar en la necesidad de hacer un estudio más a fondo sobre las asignaturas en detalle de las Ciencias Básicas.

No se puede correr la prueba de Tamhane para analizar las diferencias de medias de la probabilidad de graduarse cambiando las notas de las materias de Ciencias Básicas del estudiante, porque hay menos de tres grupos.

7.10. Análisis de los resultados

Según los resultados obtenidos, variables como el nivel de educación del responsable del estudiante no es significativa estadísticamente en las probabilidades de deserción, pero esto tiene sentido porque hoy en día los estudiantes tienen más opciones de conseguir los recursos de dinero necesarios para cubrir los costos de la educación superior.

Se analizaron las probabilidades de graduación de los estudiantes de semestres superiores a 5 y se concluyó que este valor aumenta mucho frente a los semestres de inicio. Esto puede indicar que las asignaturas de Ciencias Básicas son la causa principal, pero se requiere un análisis más profundo y en detalle de esto.

En cuanto a las edades al entrar a la carrera, como ya se mencionó, se encontró que los estudiantes más jóvenes tienen mayores probabilidades de graduarse, pero también que los estudiantes mayores terminan en menos tiempo.

Si bien las mujeres tienen mayor probabilidad de graduarse que los hombres, sus tiempos de permanencia son mucho mayores que los de estos.

8. CONCLUSIONES

Después del análisis realizado, es posible concluir que las variables que más afectan la deserción en la Escuela son el género del estudiante que sea hombre o mujer, las ciencias básicas, el estrato socioeconómico, con quién vive actualmente y el tipo de vivienda. Como se puede observar, en la Escuela Colombiana de Ingeniería no hay mucha población de los estratos 5 y 6 ni tampoco del 1, lo que predomina son los estratos 2, 3 y 4; así las cosas, los esfuerzos se deben orientar a lograr bajar la deserción en estos grupos de estudiantes.

Por ejemplo, se vio que una variable crítica es vivir solo, por lo que se debe hacer un análisis para saber si estos estudiantes se mantienen ellos mismos, o si viven solos pero las familias los apoyan, y generar estrategias de acompañamiento para procurar que no se afecte tanto su rendimiento académico. Se corrobora el resultado en el estudio de Nora¹² (Nora, 2002) en cuanto a la importancia del apoyo familiar.

En los resultados se ve que si bien las mujeres se gradúan más, también se demoran más en terminar sus estudios, por lo que sería interesante poder explorar con más detalle cuáles son las razones para esto; una causa podría ser la maternidad temprana que se viene dando en nuestra sociedad.

Los resultados obtenidos en el modelo económico corroboran el estudio realizado por Ishitani y Desjardins (Ishitani & Desjardins, 2002), en el sentido de que los estudiantes de menores ingresos tienen mayores probabilidades de graduarse.

Con los resultados obtenidos no se puede comprobar el análisis de Porto (Porto & Gresia, 2001), según el cual el nivel de educación del responsable del estudiante es un factor determinante en la deserción.

Se corrobora el estudio de Tinto (Tinto, 1975) sobre la tercera etapa de la deserción, donde se afirma que los estudiantes tienen más probabilidad de deserción debido a las asignaturas de Ciencias Básicas.

De otra parte, sería objeto de otro estudio un análisis estadístico más profundo de los resultados finales, para encontrar una prueba estadística más adecuada al tipo de datos y comportamiento de estas variables, porque con la prueba Tamhane que se aplicó se encontró que todas variables no tienen diferencias significativas en las pruebas de medias realizadas. No obstante, el modelo se comprobó con un análisis de su validez interna a través del criterio de aleatoriedad proporcional y el criterio de aleatoriedad máxima, lo que

¹² Amaury Nora realizó el estudio *The depiction of significant others in Tinto's "rites of passage": A reconceptualization of the influence of family and community in the persistence process.*

dio como resultado, que el modelo es válido estadísticamente y que los resultados no son debidos al azar (Hair et al., 1999).

9. RECOMENDACIONES

La Escuela debe documentar de manera completa y exacta toda la información de las personas admitidas, así como también preocuparse por saber, mediante una encuesta o entrevista, por qué los estudiantes desertan. Esta información es clave para seguir con estudios como este, que pueden ayudar a disminuir la deserción en la universidad, además de cumplir con un requisito de los procesos de acreditación.

Una investigación interesante, que podría ser la continuación del presente estudio, consistiría en averiguar cuáles de las asignaturas son las realmente críticas y cuál es la metodología que se usa, por ejemplo clases magistrales o no, exámenes conjuntos o no, etc.

Llama la atención que estudiantes del estrato socioeconómico 1 tengan una probabilidad más alta de graduarse. Valdría la pena hacer una investigación en la que se analizara qué motiva a estas personas a persistir a través del tiempo hasta lograr graduarse, así como también qué perfiles son los que perseveran y terminan sus estudios con éxito.

También sería interesante un análisis de nacimiento–muerte para el fenómeno de la deserción acorde con el crecimiento de la población nacional.

10. REFERENCIAS BIBLIOGRÁFICAS

- Ajzen, I. & Fishbein, M., 1977. Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, 84(5), pp.888–918.
- Astin, A. w., 1977. *Four Critical Years* S. Francisco, ed., Jossey-Bass Publishers.
- Astin, A.W. & Denson, N., 2009. Multi-Campus Studies of College Impact: Which Statistical Method is Appropriate? *Research in Higher Education*, 50(4), pp.354–367.
- Berger, J.B. & Braxton, J.M., 1998. Revising Tinto's Interactionist Theory of Student Departure Through theory elaboration. *Research in Higher Education*, 39(2), pp.103–119.
- Bessent, A. et al., 1986. An application of mathematical programming to assess productivity in the Houston Independent School District. *Journal of Management Science*, 28(0025-1909), pp.1355–1368.
- Bessent, W. & Bessent, A., 1980. Student Education Flow in a University Department Results of a Markov Analysis. *Journal of Management Science*, 10(0092-2102), p.8.
- Borden, V. & Dalphin, J., 1998. *Simulating the effect of student profile chnages on retention and graduation rates a Markov analysis.*,
- Cabrera, A.F., Nora, A. & Castañeda, M.B., 1993. College Persistence: Structural Equations Modeling Test of an Integrated Model of Student Retention. *Higher Education*, 64(2), pp.123–139.
- Castaño, E. et al., 2004. Deserción estudiantil universitaria : una aplicación de modelos de duración. *Lecturas de Economía*, (60), pp.39–65.
- Choy, S.P. et al., 2000. Transition to College: What Helps At-Risk Students and Students Whose Parents Did Not Attend College. In *Understanding the college choice of disadvantaged students*. pp. 45–63.
- DANE, 2011. Colombia. número de alumnos matriculados por nivel educativo y zona. , p.1.
- DeMaris, A., 1993. Odds versus probabilities in Logit Equations: A reply to Roncek. *Social Forces*, 71(4), pp.1057–1065.
- Donoso, S. & Schiefelbein, E., 2007. Analisis de los modelos explicativos de retencion de estudiantes en la universidad: una vision desde la desigualdad social. *Estudios pedagógicos (Valdivia)*, 33(1), pp.7–27.
- Ethington, C. a., 1990. A psychological model of student persistence. *Research in Higher Education*, 31(3), pp.279–293. Available at: <http://www.springerlink.com/index/10.1007/BF00992313>.

- Green, R., 1998. *Retention and achievement in one first year program: A multivariate analysis*.
- Hader, J.A., 2011. *William G. Spady, Agent of Change: An Oral History*. Loyola University Chicago.
- Hair, J.F. et al., 1999. *Análisis Multivariante* 5a. ed. A. Otero, ed., Madrid: Pearson Prentice Hall.
- Ishitani, T.T. & Desjardins, S.L., 2002. A Longitudinal Investigation of Dropout from College in the United States. *Journal of College Student Retention: Research, Theory and Practice*, 4(2), pp.173–201.
- John, M., Jeffrey, F. & Sullivan, A.S., 2000. The influence of active learning on the college student departure process. *Journal of Higher Education*, 71(5), pp.569–588.
- Lamb, S. et al., 2011. *School Dropout and Completion. International Comparative Studies in Theory and Policy* 1st ed. Springer, ed., London New York: Dordrecht Heidelberg.
- Leppel, K., 2005. College persistence and student attitudes toward financial success. *College Student Journal*, 39(2), pp.223–240.
- Liao, T.F., 1994. *Interpreting Probability models Logit, Probit, and other Generalized Linear Models* A. Virding, ed., California, Nueva Delhi: SAGE Publications, Inc.
- MEN, M. de E., 2008. *Deserción estudiantil en la Educación Superior en Colombia. Elementos para su diagnóstico y tratamiento.*,
- MEN, M. de E., 2010. *Jaque a la deserción*,
- Menard, S. & Houston, S., 2011. Standards for Standardized Logistic Regression Coefficients. *Social Forces*, 89(June), pp.1409–1428.
- Nora, A., 2002. The depiction of significant others in Tinto's "rites of passage": A reconceptualization of the influence of family and community in the persistence process. *College Student Journal*, 3(1), pp.41–56.
- Pascarella, E.T., 2006. How College Affects Students: Ten Directions for Future Research. *Journal of College Student Development*, 47(5), pp.508–520.
- Pascarella, E.T. & Terenzini, P.T., 1980. Predicting Freshman Persistence and Voluntary Dropout Decisions Model from a Theoretical Model. *Higher Education*, 51(1), pp.60–75.
- Perry, S.R. & Cabrera, A.F., 2000. Career Maturity and College Student Persistence. *College Student Journal*, 1(1), pp.41–58.

- Peña, M. & MEN, M. de E., 2006. Ministerio de Educación Nacional educación: visión 2019. *Ministerio de Educación MEN*, 2109, p.62.
- Porto, A. & Gresia, L. Di, 2001. *Rendimiento de estudiantes universitarios y sus determinantes*,
- Ross, S., 1996. *Stochastic Processes* 2da. ed. B. Wiley, ed., John Wiley & sons, Inc.
- SINIES, S.N. de I. de la E.S., 2012. *Matriculados en primer curso en Educación Superior - marzo 2012*,
- SPADIES, S. para la P. de la D. de la E.S., 2012. *Reporte de deserción para el sistema de Educación Superior 2012*,
- Spady, W.G., 1970. *Dropouts From Higher Education : An Interdisciplinary Review and Synthesis*,
- Spady, W.G., 1971. *Dropouts from Higher Education : Toward an Empirical Model*,
- Tinto, V., 1975. *Introducción La deserción como comportamiento individual La deserción y el carácter de las metas individuales.*,
- Tinto, V., 1993. *Leaving College: rethinking the causes and cures of student attrition*,
- Tinto, V., 2010. Limits of Theory and Practice in Student Attrition. *Journal of Higher Education*, 53(6), pp.687–700.
- Tinto, V., 2007. Research and practice of Student Retention: What next? *College Student Journal*, 8(1), pp.1–19.
- Tinto, V. & Cullen, J., 1975. Dropout from Higher Education: A Theoretical Synthesis of Recent Research. *Journal of the Educational Research*, 43(1), pp.89–125.
- Universidad Nacional de Colombia, U. & ICFES, 2002. *Estudio de la deserción en la Educación Superior en Colombia. Documento sobre Estado del Arte*,
- Vásquez, J. et al., 2003. Determinantes de la deserción estudiantil en la Universidad de Antioquía. *Centro de investigaciones económicas*, 4(1), pp.1–40.
- Winston, W.L., 2008. *Investigación de operaciones, aplicaciones y algoritmos* 4a. ed. J. Arellano, ed., Mexico D.F.: Cengage Learning.

Anexo 1 Resultados regresión logística por semestres.

Variables	1sem	2sem	3sem	4sem	5sem	6sem	7sem	8sem	9sem	10sem
Intersección SsR	,861	-,298	-,650	-,2022	,801	,100	-,489	-,2,369	-,4,782	-,2,244
Intersección SsP	1,931	,897	,372	-,1,026	1,929	1,262	,645	-,1,286	-,3,612	-,1,121
Estrato_1	,258	-,010	,983	-,280	-,577	-,099	-,1,658	1,786	1,297	-,1,930
Estrato_2	-,144	,139	,333	,562	,089	,932	-,803	2,260	1,530	1,365
Estrato_3	,414	,514	,212	,792	-,351	,833	-,1,256	2,109	1,551	2,574
Estrato_4	,146	,362	,041	-,678	-,527	,835	-,901	1,346	1,578	2,047
Estrato_5	-,131	,309	,252	,293	,988	-,012	-,491	2,318	1,586	0
Estrato_6	0	0	0	0	0	0	0	0	0	0
Sexo_1	-,137	-,336	-,166	-,155	-,147	-,692	-,273	0	-,165	0
Sexo_2	0	0	0	0	0	0	0	0	0	0
Viveactualmente_1	0	,833	0	-,2,286	0	2,329	1,266	0	0	0
Viveactualmente_2	0	,153	0	,665	0	,022	,037	0	-,423	0
Viveactualmente_3	0	,081	0	,476	0	-,102	-,183	0	-,207	0
Viveactualmente_4	0	,258	0	1,308	0	-,1,733	-,381	0	,391	0
Viveactualmente_5	0	0	0	0	0	0	0	0	0	0
Programa1_1	-,286	,267	-,010	-,333	-,395	-,375	-,180	,062	-,457	-,625
Programa1_2	-,178	,414	-,071	-,122	-,158	-,013	,950	,795	,182	,197
Programa1_3	-,115	,307	,153	,121	-,116	-,227	,301	,875	-,520	-,009
Programa1_4	-,403	,198	,105	-,027	-,314	-,076	-,410	-,390	,342	-,639
Programa1_5	0	0	0	0	0	0	0	0	0	0
Estado1_1	-,1,934	-,1,744	-,1,693	-,1,611	-,1,577	-,1,744	-,1,382	-,1,548	-,917	-,1,037
Estado1_2	-,1,586	-,1,272	-,542	-,617	-,519	-,969	-,292	-,573	-,122	,220
Estado1_3	,653	,524	,015	,041	-,137	,467	-,050	,744	,314	-,721
Estado1_4	0	0	0	0	0	0	0	0	0	0
Ne1_0	,173	0	-,175	,564	,246	-,653	-,384	0	-,653	,466
Ne1_1	-,049	0	,306	,445	-,150	-,517	,163	0	,306	-,1,097
Ne1_2	,005	0	,020	,241	-,108	-,124	,439	0	1,398	,247
Ne1_3	,125	0	-,162	,242	-,437	,057	-,108	0	1,135	-,826
Ne1_4	0	0	0	0	0	0	0	0	0	0
Re1_1	0	-,246	-,069	0	0	0	-,759	0	0	0
Re1_2	0	,085	,209	0	0	0	,257	0	0	0
Re1_3	0	0	0	0	0	0	0	0	0	0
Tipvivi1_1	-,061	-,087	-,062	-,066	-,224	-,345	-,063	-,625	0	0
Tipvivi1_2	-,009	,141	-,088	-,021	,358	-,260	,490	,761	0	0
Tipvivi1_3	,066	,235	,468	-,149	,019	,294	,368	,982	0	0
Tipvivi1_4	-,183	-,130	-,342	,154	-,481	-,458	,582	,225	0	0
Tipvivi1_5	0	0	0	0	0	0	0	0	0	0
ZEdadentrar	,154	0	0	,019	0	0	0	,044	-,150	0
ZMatricula	0	0	,057	0	0	,113	0	0	0	0
Ztotalingreso	0	0	-,114	0	,092	0	,052	-,025	,020	-,541
ZAGEO	-,082	-,026	0	0	0	0	0	-,428	0	0
ZPREC	-,123	0	0	0	0	0	0	0	0	0
ZFFIS	-,005	-,059	0	0	0	0	,196	0	0	0
ZALLI	,072	-,044	-,008	0	0	-,181	0	0	0	0
ZCALD	,007	0	-,061	0	0	0	,306	0	,217	0
ZCALI	0	0	-,040	0	0	0	0	0	0	0
ZCALV	0	0	0	-,012	-,047	0	0	0	0	0
ZECDI	0	0	0	0	-,062	,041	0	0	0	0
ZQUIM	0	0	0	0	0	-,065	0	0	0	0
ZBIOL	0	0	0	0	0	0	0	0	0	0
ZFIS1	,031	0	-,023	,023	0	0	0	0	0	0
ZFIS2	-,046	0	,092	,022	0	0	0	0	0	0
ZFIS3	0	0	0	0	-,097	,008	0	,148	0	0
ZMat	-,194	-,138	-,173	-,120	-,139	-,332	,013	-,036	,337	-,239
ZFis	-,023	-,118	,041	-,073	-,136	0	,050	0	-,012	0
ZLeng	-,025	-,058	-,123	-,210	0	0	-,197	0	,065	-,030
Naturaleza_1	0	,125	0	,068	0	0	0	0	-,522	0
Naturaleza_2	0	0	0	0	0	0	0	0	0	0
Calendario_1	-,782	-,369	,109	0	,206	0	-,525	,870	,136	-,840
Calendario_2	-,615	,015	-,112	0	-,862	0	-,351	,616	-,716	-,599
Calendario_3	0	0	0	0	0	0	0	0	0	0

Tabla 28 Betas de la regresión logística multinomial. Resumen.

Semestre 1.

Contrastes de la razón de verosimilitud

Efecto	Criterio de ajuste del modelo	Contrastes de la razón de verosimilitud		
	-2 log verosimilitud del modelo reducido	Chi-cuadrado	gl	Sig.
Intersección	10162,6	,000	0	.
Estrato	10199,4	36,810	10	,000
Sexo	10176,7	14,086	2	,001
Programa1	10218,1	55,511	8	,000
Estado1	12096,9	1934,24	6	,000
Tipviven1	10193,1	30,453	8	,000
ZEdadentrar	10191,6	28,948	2	,000
ZAGEO	10209,9	47,305	2	,000
ZPREC	10180,6	17,949	2	,000
ZFFIS	10169,7	7,049	2	,029
ZALLI	10174,2	11,589	2	,003
ZCALD	10169,9	7,326	2	,026
ZFIS1	10170,3	7,673	2	,022
ZFIS2	10165,2	2,595	2	,273
ZMat	10226,1	63,504	2	,000
ZFis	10186,5	23,874	2	,000
ZLeng	10217,1	54,543	2	,000
Calendario	10188,5	25,876	4	,000

El estadístico de chi-cuadrado es la diferencia en las -2 log verosimilitudes entre el modelo final y el modelo reducido. El modelo reducido se forma omitiendo un efecto del modelo final. La hipótesis nula es que todos los parámetros de ese efecto son 0.

a. Este modelo reducido es equivalente al modelo final ya que la omisión del efecto no incrementa los grados de libertad.

Los otros resultados se presentan en la carpeta denominada regresión logística.

Anexo 2 Operaciones con las matrices de transición de los diferentes perfiles.

Este anexo se encuentra en la carpeta denominada Matrices.

Para el siguiente nivel de educación, Técnica se tienen los siguientes resultados:

Nivel Edu	Estrato	Sexo	Vive actualmente	Responsable	Tipo de vivienda	Edad	Ingreso
Técnica	3	Hombre	Con padres	Madre	Propia y pagada	18	Medio

Id	sem1	sem2	sem3	sem4	sem5	sem6	sem7	sem8	sem9	sem10	Graduan	Retiran	Suma
Sem1	0,186320472	0,66803932	0	0	0	0	0	0	0	0	0	0,1456402	1
Sem2	0	0,16986367	0,72905502	0	0	0	0	0	0	0	0	0,10108131	1
Sem3	0	0	0,11913409	0,79679619	0	0	0	0	0	0	0	0,08406972	1
Sem4	0	0	0	0,12731029	0,77662885	0	0	0	0	0	0	0,09606086	1
Sem5	0	0	0	0	0,11521481	0,81729211	0	0	0	0	0	0,06749307	1
Sem6	0	0	0	0	0	0,18238923	0,6988425	0	0	0	0	0,11876826	1
Sem7	0	0	0	0	0	0	0,07648646	0,88240635	0	0	0	0,04110719	1
Sem8	0	0	0	0	0	0	0	0,12551421	0,79352097	0	0	0,08096482	1
Sem9	0	0	0	0	0	0	0	0	0,06518198	0,90232249	0	0,03249553	1
Sem10	0	0	0	0	0	0	0	0	0	0,03440606	0,94808674	0,0175072	1
Graduan	0	0	0	0	0	0	0	0	0	0	0	1	1
Retiran	0	0	0	0	0	0	0	0	0	0	0	1	1

Tabla No. 29 Matrices de transición perfil 3.

$M=(I-Q)^{-1}$	1	2	3	4	5	6	7	8	9	10	Tiempo de permanencia
1	1,22899	0,98901	0,81856	0,74737	0,65601	0,65576	0,496226	0,50072	0,42504	0,397186	6,91486
2	0	1,20462	0,99701	0,91031	0,79903	0,79872	0,604409	0,60988	0,5177	0,483777	
3	0	0	1,13525	1,03652	0,90982	0,90946	0,688208	0,69444	0,58948	0,550851	
4	0	0	0	1,14588	1,00581	1,00542	0,760821	0,76771	0,65167	0,608972	
5	0	0	0	0	1,13022	1,12978	0,854927	0,86267	0,73228	0,684295	
6	0	0	0	0	0	1,22308	0,925528	0,93391	0,79275	0,740805	
7	0	0	0	0	0	0	1,082821	1,09263	0,92748	0,866705	
8	0	0	0	0	0	0	0	1,14353	0,97069	0,90708	
9	0	0	0	0	0	0	0	0	1,06973	0,999632	
10	0	0	0	0	0	0	0	0	0	1,035632	

M*R = Nivel de educación Técnica		
	Gradúan	Retiran
1	0,37657	0,62343
2	0,45866	0,54134
3	0,52225	0,47775
4	0,57736	0,42264
5	0,64877	0,35123
6	0,70235	0,29765
7	0,82171	0,17829
8	0,85999	0,14001
9	0,94774	0,05226
10	0,98187	0,01813

Para el siguiente nivel de educación, Universitaria se tienen los siguientes resultados:

Nivel Edu	Estrato	Sexo	Vive actualmente	Responsable	Tipo de vivienda	Edad	Ingreso
Universitaria	3	Hombre	Con padres	Madre	Propia y pagada	18	Medio

Id	sem1	sem2	sem3	sem4	sem5	sem6	sem7	sem8	sem9	sem10	Graduan	Retiran	Suma
Sem1	0,165196166	0,713985346	0	0	0	0	0	0	0	0	0	0,12081849	1
Sem2	0	0,165758709	0,73661508	0	0	0	0	0	0	0	0	0,09762621	1
Sem3	0	0	0,13008829	0,77560357	0	0	0	0	0	0	0	0,09430814	1
Sem4	0	0	0	0,10758504	0,81506579	0	0	0	0	0	0	0,07734918	1
Sem5	0	0	0	0	0,15458608	0,74623406	0	0	0	0	0	0,09917986	1
Sem6	0	0	0	0	0	0,17382887	0,71563302	0	0	0	0	0,11053811	1
Sem7	0	0	0	0	0	0	0,08502955	0,86854237	0	0	0	0,04642808	1
Sem8	0	0	0	0	0	0	0	0,14804111	0,75106453	0	0	0,10089436	1
Sem9	0	0	0	0	0	0	0	0	0,02112183	0,9690735	0	0,00980466	1
Sem10	0	0	0	0	0	0	0	0	0	0,07728974	0,88035637	0,04235389	1
Graduan	0	0	0	0	0	0	0	0	0	0	1	0	1
Retiran	0	0	0	0	0	0	0	0	0	0	0	1	1

M*R = Nivel de educación Universitaria o más

	Gradúan	Retiran
1	0,37165	0,62835
2	0,43454	0,56546
3	0,49213	0,50787
4	0,55197	0,44803
5	0,60435	0,39565
6	0,68467	0,31533
7	0,79043	0,20957
8	0,83268	0,16732
9	0,94454	0,05546
10	0,95410	0,04590

El detalle de las otros perfiles se encuentra en el Anexo No. 2 Operaciones con las matrices de transición de los perfiles analizados, a continuación se presenta un resumen con las conclusiones y los análisis de los resultados obtenidos.

Anexo 3 Programa en Visual Basic Studio 2008.

```

Public Class Form1
    Private Sub Form1_Load(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles MyBase.Load
        'TODO: esta línea de código carga datos en la tabla 'InformalDataSet9.betas'
Puede moverla o quitarla según sea necesario.
        Me.BetasTableAdapter4.Fill(Me.InformalDataSet9.betas)
        'TODO: esta línea de código carga datos en la tabla 'InformalDataSet8.betas'
Puede moverla o quitarla según sea necesario.
        Me.BetasTableAdapter3.Fill(Me.InformalDataSet8.betas)
        'TODO: esta línea de código carga datos en la tabla 'InformalDataSet7.betas'
Puede moverla o quitarla según sea necesario.
        Me.BetasTableAdapter2.Fill(Me.InformalDataSet7.betas)
        'TODO: esta línea de código carga datos en la tabla
'InformalDataSet6.variables' Puede moverla o quitarla según sea necesario.
        Me.VariablesTableAdapter1.Fill(Me.InformalDataSet6.variables)
        'TODO: esta línea de código carga datos en la tabla 'InformalDataSet5.dato'
Puede moverla o quitarla según sea necesario.
        Me.DatoTableAdapter1.Fill(Me.InformalDataSet5.dato)
        'TODO: esta línea de código carga datos en la tabla 'InformalDataSet4.dato'
Puede moverla o quitarla según sea necesario.
        Me.DatoTableAdapter.Fill(Me.InformalDataSet4.dato)
        'TODO: esta línea de código carga datos en la tabla 'InformalDataSet3.matriz'
Puede moverla o quitarla según sea necesario.
        Me.MatrizTableAdapter3.Fill(Me.InformalDataSet3.matriz)
        'TODO: esta línea de código carga datos en la tabla 'InformalDataSet2.matriz'
Puede moverla o quitarla según sea necesario.
        Me.MatrizTableAdapter2.Fill(Me.InformalDataSet2.matriz)
        'TODO: esta línea de código carga datos en la tabla 'InformalDataSet1.matriz'
Puede moverla o quitarla según sea necesario.
        Me.MatrizTableAdapter1.Fill(Me.InformalDataSet1.matriz)

        'TODO: esta línea de código carga datos en la tabla 'InformalDataSet.matriz'
Puede moverla o quitarla según sea necesario.
        Me.MatrizTableAdapter.Fill(Me.InformalDataSet.matriz)
        'TODO: esta línea de código carga datos en la tabla
'InformalDataSet.variables' Puede moverla o quitarla según sea necesario.
        Me.VariablesTableAdapter.Fill(Me.InformalDataSet.variables)
        'TODO: esta línea de código carga datos en la tabla 'InformalDataSet.betas'
Puede moverla o quitarla según sea necesario.
        Me.BetasTableAdapter2.Fill(Me.InformalDataSet7.betas)
        Me.ListView1.Clear()
        Me.ListView1.Columns.Add("Sem 1", 100, HorizontalAlignment.Left)
        Me.ListView1.Columns.Add("Sem 2", 100, HorizontalAlignment.Left)
        Me.ListView1.Columns.Add("Sem 3", 100, HorizontalAlignment.Left)
        Me.ListView1.Columns.Add("Sem 4", 100, HorizontalAlignment.Left)
        Me.ListView1.Columns.Add("Sem 5", 100, HorizontalAlignment.Left)
        Me.ListView1.Columns.Add("Sem 6", 100, HorizontalAlignment.Left)
        Me.ListView1.Columns.Add("Sem 7", 100, HorizontalAlignment.Left)
        Me.ListView1.Columns.Add("Sem 8", 100, HorizontalAlignment.Left)
        Me.ListView1.Columns.Add("Sem 9", 100, HorizontalAlignment.Left)
        Me.ListView1.Columns.Add("Sem 10", 100, HorizontalAlignment.Left)
        Me.ListView1.Columns.Add("Graduan", 100, HorizontalAlignment.Left)
        Me.ListView1.Columns.Add("Retiran", 100, HorizontalAlignment.Left)
        Me.ListView1.Columns.Add("Suma", 100, HorizontalAlignment.Left)
        Me.ListView1.GridLines = True
        Me.ListView1.View = View.Details
    End Sub
    Private Sub Button1_Click(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles Button1.Click
        Dim alfals(3) As Double
        Dim alfa2s(3) As Double

```

```

Dim alfa3s(3) As Double
Dim alfa4s(3) As Double
Dim alfa5s(3) As Double
Dim alfa6s(3) As Double
Dim alfa7s(3) As Double
Dim alfa8s(3) As Double
Dim alfa9s(3) As Double
Dim alfa10s(3) As Double
Dim betaxsem1 As Double
Dim betaxsem2 As Double
Dim betaxsem3 As Double
Dim betaxsem4 As Double
Dim betaxsem5 As Double
Dim betaxsem6 As Double
Dim betaxsem7 As Double
Dim betaxsem8 As Double
Dim betaxsem9 As Double
Dim betaxsem10 As Double
Dim SsD1s, SsP1s, SsA1s As Double
Dim SsD2s, SsP2s, SsA2s As Double
Dim SsD3s, SsP3s, SsA3s As Double
Dim SsD4s, SsP4s, SsA4s As Double
Dim SsD5s, SsP5s, SsA5s As Double
Dim SsD6s, SsP6s, SsA6s As Double
Dim SsD7s, SsP7s, SsA7s As Double
Dim SsD8s, SsP8s, SsA8s As Double
Dim SsD9s, SsP9s, SsA9s As Double
Dim SsD10s, SsP10s, SsA10s As Double
Dim temp2, temp3, suma As Double
Dim prob1, prob2, prob3 As Double
Dim nom1, nom2, nom As Double
Dim probssr, probssp As Double
Dim k, l, b, v, m As Integer
Dim sa As String
Dim i, j As Integer
Dim rowsbeta, rowsmatriz, colvar As DataRowCollection
Dim fila As ListViewItem

v = Me.VariablesTableAdapter1.GetData().Columns.Count
colvar = Me.VariablesTableAdapter1.GetData().Rows

b = Me.BetasTableAdapter4.GetData().Rows.Count
rowsbeta = Me.BetasTableAdapter4.GetData().Rows

nom2 = rowsbeta.Item(0).Item(0)

m = Me.MatrizTableAdapter.GetData().Rows.Count
rowsmatriz = Me.MatrizTableAdapter.GetData().Rows

'inicia los alfas
For nom = 0 To 3
    alfa1s(nom) = 0
    alfa2s(nom) = 0
    alfa3s(nom) = 0
    alfa4s(nom) = 0
    alfa5s(nom) = 0
    alfa6s(nom) = 0
    alfa7s(nom) = 0
    alfa8s(nom) = 0
    alfa9s(nom) = 0
    alfa10s(nom) = 0
Next nom
'inicializa la matriz de respuestas

```

```

For i = 0 To 11
  For j = 0 To 12
    matriz(i, j) = 0
  Next
Next
matriz(10, 10) = 1
matriz(11, 11) = 1
'multiplica vector variables de cada semestre por cada uno de los betas

i = 4
j = 4
While i <= b - 1
  While j <= v - 1
    betaxsem1 = betaxsem1 + (rowsbeta.Item(i).Item(2) *
colvar.Item(0).Item(j))
    Me.TextBox2.Text = Str(i)
    sa = rowsbeta.Item(i).Item(1)
    i = i + 1
    j = j + 1
  End While
End While
i = 4
j = 4
While i <= b - 1
  While j <= v - 1
    betaxsem2 = betaxsem2 + (rowsbeta.Item(i).Item(3) *
colvar.Item(1).Item(j))
    i = i + 1
    j = j + 1
  End While
End While
i = 4
j = 4
While i <= b - 1
  While j <= v - 1
    betaxsem3 = betaxsem3 + (rowsbeta.Item(i).Item(4) *
colvar.Item(2).Item(j))
    i = i + 1
    j = j + 1
  End While
End While
i = 4
j = 4
While i <= b - 1
  While j <= v - 1
    betaxsem4 = betaxsem4 + (rowsbeta.Item(i).Item(5) *
colvar.Item(3).Item(j))
    i = i + 1
    j = j + 1
  End While
End While
i = 4
j = 4
While i <= b - 1
  While j <= v - 1
    betaxsem5 = betaxsem5 + (rowsbeta.Item(i).Item(6) *
colvar.Item(4).Item(j))
    i = i + 1
    j = j + 1
  End While
End While
i = 4
j = 4

```

```

While i <= b - 1
  While j <= v - 1
    betaxsem6 = betaxsem6 + (rowsbeta.Item(i).Item(7) *
colvar.Item(5).Item(j))
    i = i + 1
    j = j + 1
  End While
End While
i = 4
j = 4
While i <= b - 1
  While j <= v - 1
    betaxsem7 = betaxsem7 + (rowsbeta.Item(i).Item(8) *
colvar.Item(6).Item(j))
    i = i + 1
    j = j + 1
  End While
End While
i = 4
j = 4
While i <= b - 1
  While j <= v - 1
    betaxsem8 = betaxsem8 + (rowsbeta.Item(i).Item(9) *
colvar.Item(7).Item(j))
    i = i + 1
    j = j + 1
  End While
End While
i = 4
j = 4
While i <= b - 1
  While j <= v - 1
    betaxsem9 = betaxsem9 + (rowsbeta.Item(i).Item(10) *
colvar.Item(8).Item(j))
    i = i + 1
    j = j + 1
  End While
End While
i = 4
j = 4
While i <= b - 1
  While j <= v - 1
    betaxsem10 = betaxsem10 + (rowsbeta.Item(i).Item(11) *
colvar.Item(9).Item(j))
    i = i + 1
    j = j + 1
  End While
End While
'calculo de los alfas

alfas(0) = rowsbeta.Item(0).Item(2)
alfas(1) = rowsbeta.Item(1).Item(2)
alfas(2) = rowsbeta.Item(2).Item(2)
alfas(3) = rowsbeta.Item(3).Item(2)

alfa2s(0) = rowsbeta.Item(0).Item(3)
alfa2s(1) = rowsbeta.Item(1).Item(3)
alfa2s(2) = rowsbeta.Item(2).Item(3)
alfa2s(3) = rowsbeta.Item(3).Item(3)

alfa3s(0) = rowsbeta.Item(0).Item(4)
alfa3s(1) = rowsbeta.Item(1).Item(4)
alfa3s(2) = rowsbeta.Item(2).Item(4)

```

```

alfa3s(3) = rowbeta.Item(3).Item(4)

alfa4s(0) = rowbeta.Item(0).Item(5)
alfa4s(1) = rowbeta.Item(1).Item(5)
alfa4s(2) = rowbeta.Item(2).Item(5)
alfa4s(3) = rowbeta.Item(3).Item(5)

alfa5s(0) = rowbeta.Item(0).Item(6)
alfa5s(1) = rowbeta.Item(1).Item(6)
alfa5s(2) = rowbeta.Item(2).Item(6)
alfa5s(3) = rowbeta.Item(3).Item(6)

alfa6s(0) = rowbeta.Item(0).Item(7)
alfa6s(1) = rowbeta.Item(1).Item(7)
alfa6s(2) = rowbeta.Item(2).Item(7)
alfa6s(3) = rowbeta.Item(3).Item(7)

alfa7s(0) = rowbeta.Item(0).Item(8)
alfa7s(1) = rowbeta.Item(1).Item(8)
alfa7s(2) = rowbeta.Item(2).Item(8)
alfa7s(3) = rowbeta.Item(3).Item(8)

alfa8s(0) = rowbeta.Item(0).Item(9)
alfa8s(1) = rowbeta.Item(1).Item(9)
alfa8s(2) = rowbeta.Item(2).Item(9)
alfa8s(3) = rowbeta.Item(3).Item(9)

alfa9s(0) = rowbeta.Item(0).Item(10)
alfa9s(1) = rowbeta.Item(1).Item(10)
alfa9s(2) = rowbeta.Item(2).Item(10)
alfa9s(3) = rowbeta.Item(3).Item(10)

alfa10s(0) = rowbeta.Item(0).Item(11)
alfa10s(1) = rowbeta.Item(1).Item(11)
alfa10s(2) = rowbeta.Item(2).Item(11)
alfa10s(3) = rowbeta.Item(3).Item(11)

```

'calculo de la función

```

SsD1s = (alfa1s(1) + betaxsem1)
SsP1s = (alfa1s(2) + betaxsem1)
SsA1s = (alfa1s(3) + betaxsem1)

SsD2s = (alfa2s(1) + betaxsem2)
SsP2s = (alfa2s(2) + betaxsem2)
SsA2s = (alfa2s(3) + betaxsem2)

SsD3s = (alfa3s(1) + betaxsem3)
SsP3s = (alfa3s(2) + betaxsem3)
SsA3s = (alfa3s(3) + betaxsem3)

SsD4s = (alfa4s(1) + betaxsem4)
SsP4s = (alfa4s(2) + betaxsem4)
SsA4s = (alfa4s(3) + betaxsem4)

SsD5s = (alfa5s(1) + betaxsem5)
SsP5s = (alfa5s(2) + betaxsem5)
SsA5s = (alfa5s(3) + betaxsem5)

SsD6s = (alfa6s(1) + betaxsem6)
SsP6s = (alfa6s(2) + betaxsem6)
SsA6s = (alfa6s(3) + betaxsem6)

```

```

SsD7s = (alfa7s(1) + betaxsem7)
SsP7s = (alfa7s(2) + betaxsem7)
SsA7s = (alfa7s(3) + betaxsem7)

SsD8s = (alfa8s(1) + betaxsem8)
SsP8s = (alfa8s(2) + betaxsem8)
SsA8s = (alfa8s(3) + betaxsem8)

SsD9s = (alfa9s(1) + betaxsem9)
SsP9s = (alfa9s(2) + betaxsem9)
SsA9s = (alfa9s(3) + betaxsem9)

SsD10s = (alfa10s(1) + betaxsem10)
SsP10s = (alfa10s(2) + betaxsem10)
SsA10s = (alfa10s(3) + betaxsem10)
'inicializa la matriz de respuestas
For i = 0 To 11
    For j = 0 To 12
        matriz(i, j) = 0
    Next
Next
matriz(10, 10) = 1
matriz(11, 11) = 1
'1semestre
matriz(0, 11) = 1 / (1 + Math.Exp(-SsD1s))
matriz(0, 0) = (1 / (1 + Math.Exp(-SsP1s))) - matriz(0, 11)
matriz(0, 1) = 1 - (1 / (1 + Math.Exp(-SsP1s)))

'2semestre
matriz(1, 11) = 1 / (1 + Math.Exp(-SsD2s))
matriz(1, 1) = 1 / (1 + Math.Exp(-SsP2s)) - matriz(1, 11)
matriz(1, 2) = 1 - (1 / (1 + Math.Exp(-SsP2s)))

'3semestre
matriz(2, 11) = 1 / (1 + Math.Exp(-SsD3s))
matriz(2, 2) = (1 / (1 + Math.Exp(-SsP3s))) - matriz(2, 11)
matriz(2, 3) = 1 - (1 / (1 + Math.Exp(-SsP3s)))

'4semestre
matriz(3, 11) = 1 / (1 + Math.Exp(-SsD4s))
matriz(3, 3) = (1 / (1 + Math.Exp(-SsP4s))) - matriz(3, 11)
matriz(3, 4) = 1 - (1 / (1 + Math.Exp(-SsP4s)))

'5semestre
matriz(4, 11) = 1 / (1 + Math.Exp(-SsD5s))
matriz(4, 4) = (1 / (1 + Math.Exp(-SsP5s))) - matriz(4, 11)
matriz(4, 5) = 1 - (1 / (1 + Math.Exp(-SsP5s)))

'6semestre
matriz(5, 11) = 1 / (1 + Math.Exp(-SsD6s))
matriz(5, 5) = (1 / (1 + Math.Exp(-SsP6s))) - matriz(5, 11)
matriz(5, 6) = 1 - (1 / (1 + Math.Exp(-SsP6s)))

'7semestre
matriz(6, 11) = 1 / (1 + Math.Exp(-SsD7s))
matriz(6, 6) = (1 / (1 + Math.Exp(-SsP7s))) - matriz(6, 11)
matriz(6, 7) = 1 - (1 / (1 + Math.Exp(-SsP7s)))

'8semestre
matriz(7, 11) = 1 / (1 + Math.Exp(-SsD8s))
matriz(7, 7) = (1 / (1 + Math.Exp(-SsP8s))) - matriz(7, 11)
matriz(7, 8) = 1 - (1 / (1 + Math.Exp(-SsP8s)))

```

```

'9semestre
matriz(8, 11) = 1 / (1 + Math.Exp(-SsD9s))
matriz(8, 8) = (1 / (1 + Math.Exp(-SsP9s))) - matriz(8, 11)
matriz(8, 9) = 1 - (1 / (1 + Math.Exp(-SsP9s)))

'10semestre
matriz(9, 11) = 1 / (1 + Math.Exp(-SsD10s))
matriz(9, 9) = (1 / (1 + Math.Exp(-SsP10s))) - matriz(9, 11)
matriz(9, 10) = 1 - (1 / (1 + Math.Exp(-SsP10s)))

prob1 = SsD10s
prob2 = SsP10s
probssr = Math.Exp(-SsD10s)
probssp = Math.Exp(-SsP10s)
prob3 = 1 / (1 + Math.Exp(-SsD10s))
nom1 = 1 / (1 + Math.Exp(-SsP10s))

'rutina para mostrar la probabilidades...
For l = 0 To 11
    suma = 0
    temp2 = matriz(l, 0)
    fila = New ListViewItem(temp2)
    temp3 = 1
    k = 1
    suma = suma + matriz(l, 0)
    For k = 1 To 11
        fila.SubItems.Add(matriz(l, k))
        suma = suma + matriz(l, k)
    Next
    matriz(l, 12) = suma
    fila.SubItems.Add(matriz(l, 12))
    ListView1.Items.AddRange(New ListViewItem() {fila})
    Me.ListView1.View = View.Details
Next
Me.ListView1.View = View.Details
End Sub

Private Sub Button2_Click(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles Button2.Click
    'rutina para grabar las probabilidades
    Dim demu As Double
    Dim mona As Integer

    demu = matriz(0, 0)

    For mona = 0 To 11
        Me.DatoTableAdapter1.InsertQuery(matriz(mona, 0), matriz(mona, 1),
matriz(mona, 2), matriz(mona, 3), matriz(mona, 4), matriz(mona, 5), matriz(mona,
6), matriz(mona, 7), matriz(mona, 8), matriz(mona, 9), matriz(mona, 10),
matriz(mona, 11), matriz(mona, 12))
        Me.InformalDataSet5.dato.AcceptChanges()
        Me.DatoTableAdapter1.Fill(Me.InformalDataSet5.dato)
    Next
    demu = demu
    Me.TextBox1.Text = "Ya grabó"
End Sub
End Class

```

Anexo 4 Descripción de los archivos iniciales.

Código	Información del estudiante
id_est	Código estudiante
per_ing	Periodo de ingreso
sexo	Género del estudiante
fec_nac	Fecha de Nacimiento
estrato	Estrato del estudiante
nom_dpto	Nombre programa de ingeniería
cod_plan	Código del plan de estudios
Mat	Nota del Icfes en matemáticas
Fis	Nota del Icfes en física
Leng	Nota del Icfes en lenguaje
Fil	Nota del Icfes en filosofía
id_col	Código del colegio
nom_col	Nombre del colegio
cal_col	Calendario colegio
naturaleza	Naturaleza del colegio
fec_grado	Fecha de grado del colegio
base_liq	Valor de la Matrícula
dest	Estado del estudiante
vive_act	Con quién vive actualmente el estudiante
sem_est	Último semestre reportado en el sistema

Código	Información académica
idest	Código estudiante
codasig	Código de la asignatura
nomasig	Nombre de la asignatura
nota	Nota
clave	Estado de la asignatura, cancelada, normal
cred	Número de créditos de la asignatura
prom	Flag que dice si la asignatura se computa o no
perAcad	Periodo cuando se cursa la asignatura

Código	Información financiera
idest	Código estudiante
Responsable 1	Quién es el responsable
NE	Nivel de educación
Profesión	Qué profesión tiene
Ingreso(x1000)	Ingresos
Renta Líquida(x1000)	Renta
estrato	Estrato socioeconómico
Tipo de Vivienda	Tipo de vivienda
Responsable 2	Quién es el responsable 2
Nivel Educación	Nivel de educación
Profesión	Qué profesión tiene
Ingreso(x1000)	Ingresos
Renta Líquida(x1000)	Renta
Estrato	Estrato socioeconómico
Tipo de Vivienda	Tipo de vivienda

Anexo 5 Reducción y limpieza de los datos.

- Se eliminan las carreras que no son ingenierías, e Ingeniería Mecánica, porque apenas en el año 2010 empezó el programa y no hay una cohorte completa.
- Se eliminan los registros que tienen en estado el valor de convenio, porque son estudiantes de colegio que no están matriculados formalmente. En caso de que continúen, se les asigna un nuevo número de identificación; además, estas personas no pagan y no se consideran estudiantes regulares.
- El archivo suministrado por el área financiera de la Escuela contenía registros duplicados, por lo que se decidió eliminar los datos repetidos. Esto tomó ocho horas de trabajo.
- Se elimina la variable nombre de colegio, porque no se va a usar.
- Se borran columnas en blanco y totales al final de control de archivo.
- Se eliminan asignaturas que tienen nota de 60 y de 90 porque son deporte e inglés, que no se tienen en cuenta ni para el promedio ni para el pécsum de la carrera; son como adicionales de Bienestar.
- Se eliminan las materias canceladas porque no tienen efecto en el promedio, ya que no suman ni tienen ningún cálculo.
- Se elimina la variable de plan de estudios porque no se va a usar, al igual que las profesiones de los responsables; esta última se recodificó, pero en vista de que no salió significativa se quitó definitivamente.
- Se eliminan las asignaturas de homologación pero después se quita toda la persona porque causa ruido, ya que al no entrar al inicio del primer semestre daña los cálculos del programa, porque habría que entrarlo a algún semestre que no es el primero, incumpliendo así con las condiciones de las propiedades de las cadenas de Markov absorbentes.
- Se elimina la variable fecha de grado porque tiene demasiados valores ausentes y no salió significativa en los modelos iniciales.

Anexo 6 Recodificación de variables.

- Se arreglan códigos que estaban errados. Por ejemplo, calendario A estaba con doble A (AA).
- Se arreglan algunas fechas que quedaron en formato numérico y se convierten a formato de fecha.
- Se arreglan algunas fechas de nacimiento que tienen fecha de grado, entonces con base en esta se corrigen las de nacimiento.
- Se crea una variable que se denomina excluido, la cual toma el valor de 1 si el estudiante se retira definitivamente. Los conceptos son cancelado, cancelado-prueba, excluido sin seguimiento, excluido, expulsado, no pago, no pago prueba, tres materias, adm. seguimiento, anulado, excluido 1er. sem, no admit. readmisión, no admit. reingreso, no admit. reintegro, no matric. readmisión, no pago 2da. prueba.
- Se recodifican los programas de ingeniería, como se puede observar en la tabla de variables administrativas.
- Se recodifica el calendario del colegio, la naturaleza del colegio, si es privada u oficial, el sexo, con quién vive actualmente, el nivel de educación, el tipo de responsable y el tipo de vivienda. Esto se puede observar en detalle en la tabla de variables independientes, en la sección anterior.
- Se recodifican todas las asignaturas, que son más de 2000, para dejarlas agrupadas según su tema –ciencias básicas, básicas de ingeniería, ingeniería aplicada, económico-administrativas, humanísticas, electivas–; también se identifican las materias de ciencias básicas más importantes para un análisis más detallado, por lo que se colocan como variables dentro del modelo para saber cuáles son las más críticas. Este proceso duró doce horas.
- Se recodifica el periodo intersemestral, que está como una I y se coloca 3; así quedan 1, primer semestre, y 2, segundo semestre.
- Se calcula la edad del estudiante y se borra la fecha de nacimiento porque no salió significativa en los modelos iniciales.
- Se recodifican todas las variables escalares con los valores tipificados, para suavizar y mejorar la dimensionalidad de los datos, es decir, que no sean valores excesivamente grandes que afecten los valores de los resultados.

Anexo 7 Deducción detallada del modelo de la regresión logística multinomial.

El modelo utilizado para el cálculo de las probabilidades fue (Liao 1994)¹³

$$y^* = \sum_{k=1}^K \beta_k x_k + \epsilon \quad [1]$$

En la práctica y^* es inobservado y ϵ es simétricamente distribuida con media cero y tiene una distribución acumulativa (CDF)¹⁴ definida como $F(\epsilon)$.

Lo que realmente se observa es una variable ficticia y , una realización de un proceso binomial, defina por

$$y = \begin{cases} 0 & \text{si } y^* > 0, \\ 1 & \text{en otro caso.} \end{cases}$$

En el modelo formulado en la ecuación [1], el termino sumado $\beta_k x_k$ es $-E(y|x_1, \dots, x_k)$ como en el caso lineal, pero $E(y^* | x_1, \dots, x_k)$.

De estas relaciones se tiene:

$$Prob(y = 1) = Prob \left[\sum_{k=1}^K \beta_k x_k + \epsilon > 0 \right]$$

$$Prob(y = 1) = Prob \left[\epsilon > - \sum_{k=1}^K \beta_k x_k \right]$$

$$Prob(y = 1) = 1 - F \left[- \sum_{k=1}^K \beta_k x_k \right],$$

Donde F es una CDF la función de distribución acumulativa de ϵ .

Se puede ver que η , como el predictor lineal generalizado, como el componente sistemático en y^* , y la ϵ como el componente aleatorio en y^* .

La forma funcional de F depende de la distribución o más bien de la suposición hecha sobre la distribución de ϵ en

$$y^* = \sum_{k=1}^K \beta_k x_k + \epsilon$$

¹³ Este modelo estadístico fué tomado del libro Interpreting Probability Models Logit, Probit, and Other Generalized linear Models de Tim Futing Liao.

¹⁴ CDF: Cumulative Distribution Function, Función de distribución acumulativa.

Obviamente la distribución de ϵ determina la función de enlace del modelo lineal generalizado.

Cuando se asume que el componente aleatorio de la respuesta en los datos, sigue una distribución binomial, se puede además asumir una distribución logística para ϵ . Además los modelos Logit se pueden aplicar a los datos y la función de enlace se convierte en Logit:

$$\eta = \log \left[\frac{\mu}{(1 - \mu)} \right]$$

Aplicando esta función se especifica un modelo Logit que toma una variable de salida binaria. El modelo Logit generalmente toma dos formas puede ser expresado en los términos de Logit o puede ser expresado en los términos de probabilidad del acontecimiento. Cuando se expresa en la forma Logit el modelo se especifica así:

$$\log \left[\frac{P(y = 1)}{1 - P(y = 1)} \right] = \sum_{k=1}^K \beta_k x_k$$

Porque ahora se modela la probabilidad que el acontecimiento A ocurra o la $Prob(y = 1)$, la μ se convierte en la probabilidad esperada de y igual a 1.

$$\mu = Prob(y = 1)$$

Usando esta ecuación

$$Prob(y = 1) = 1 - F \left[- \sum_{k=1}^K \beta_k x_k \right]$$

La ecuación

$$\log \left[\frac{P(y = 1)}{1 - P(y = 1)} \right] = \sum_{k=1}^K \beta_k x_k$$

se puede transformar en una especificación del modelo Logit de probabilidad de eventos al reemplazar la CDF general, F , por una CDF, L , específica que represente la distribución logística

$$Prob(y = 1) = 1 - L \left[- \sum_{k=1}^K \beta_k x_k \right] = L \left[\sum_{k=1}^K \beta_k x_k \right] = \frac{e^{\sum_{k=1}^K \beta_k x_k}}{1 + e^{\sum_{k=1}^K \beta_k x_k}}$$

Esto representa la probabilidad de ocurrencia de un evento. La probabilidad de un no evento es justo 1 menos la probabilidad del evento o

$$Prob(y = 0) = L \left[- \sum_{k=1}^K \beta_k x_k \right] = \frac{e^{-\sum_{k=1}^K \beta_k x_k}}{1 + e^{-\sum_{k=1}^K \beta_k x_k}} = \frac{1}{1 + e^{\sum_{k=1}^K \beta_k x_k}}$$

Modelo Logit ordinal

Es una extensión natural del modelo binario, construido alrededor de regresión latente similar al modelo binario Logit

$$y^* = \sum_{k=1}^K \beta_k x_k + \epsilon$$

Como antes y^* es inobservado y así puede ser pensado como la tendencia subyacente de un fenómeno observado. Se asume que ϵ sigue cierta distribución simétrica con media cero como la distribución normal o logística, lo que se observa es

$$y = 1 \text{ si } y^* \leq \mu_1 (= 0),$$

$$y = 2 \text{ si } \mu_1 \leq y^* \leq \mu_2,$$

$$y = 3 \text{ si } \mu_2 \leq y^* \leq \mu_3,$$

·

·

·

$$y = J \text{ si } \mu_{j-1} \leq y^* ,$$

Donde y es observado en J número de categorías ordenadas y μ_s son parámetros desconocidos que separan las categorías adyacentes que son estimadas con β_s . En general

$$Prob(y = j) = F \left[\mu_j - \sum_{k=1}^K \beta_k x_k \right] - F \left[\mu_{j-1} - \sum_{k=1}^K \beta_k x_k \right]$$

Esta ecuación da la forma general de que la y observada caiga en la categoría j y las μ_s y las β_s deben ser estimados con un modelo Logit ordinal.

Aquí se usa la función F de distribución general acumulativa, en lugar de especificar de forma particular una distribución (que podría ser logística o normal). Para que las probabilidades sean positivas se debe cumplir:

$$0 < \mu_2 < \mu_3 < \dots < \mu_{j-1}$$

El primer parámetro μ_1 del umbral típicamente se normaliza a cero para tener un parámetro menos para estimar, esto es factible porque la escala es arbitraria y puede comenzar y terminar con cualquier valor. Sin esta normalización habría $j - 1$ números de μ s para estimar porque el número de umbrales es siempre uno menos más pequeño que el número de categorías. Con esta normalización ($\mu_1 = 0$) habría $j - 2$ números de μ s para estimar. Como antes la función Logit o la función Probit puede ser usada. El modelo ordinal Logit es:

$$\log \left[\frac{P(y \leq j|x)}{1 - P(y \leq j|x)} \right] = \mu_j - \sum_{k=1}^K \beta_k x_k, \quad j = 1, 2, \dots, j - 1$$

La única diferencia entre el Logit ordinal y el binario es que el modelo de resultado ordinal permite una secuencia de probabilidades de log-odds o Logit especificado con la misma β s y x s pero diferente μ s.

También el lado izquierdo de la ecuación se llama el Logit acumulativo, con una alternativa de proporciones se puede expresar la misma relación Logit en probabilidades, esta forma se conoce como la regresión logística:

$$Prob(y \leq j) = Prob(y^* \leq \mu_j) = \frac{e^{\mu_j - \sum_{k=1}^K \beta_k x_k}}{1 + e^{\mu_j - \sum_{k=1}^K \beta_k x_k}}$$

Esto da la distribución acumulativa logística se puede expresar el Logit en términos de probabilidad

$$\begin{aligned} Prob(y = 1) &= L \left[- \sum_{k=1}^K \beta_k x_k \right], \\ Prob(y = 2) &= L \left[\mu_2 - \sum_{k=1}^K \beta_k x_k \right] - L \left[- \sum_{k=1}^K \beta_k x_k \right], \\ Prob(y = 3) &= L \left[\mu_3 - \sum_{k=1}^K \beta_k x_k \right] - L \left[\mu_2 - \sum_{k=1}^K \beta_k x_k \right] \\ &\vdots \\ Prob(y = J) &= 1 - L \left[\mu_{j-1} - \sum_{k=1}^K \beta_k x_k \right] \end{aligned}$$

En el SPSS® se estiman modelos ordinales que se usan las 3 funciones incluyendo el Logit y el Probit porque escribe la función como:

$$\log \left[\frac{\text{Prob}(y \leq j|x)}{1 - \text{Prob}(y \leq j|x)} \right] = \alpha_j + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k$$

Donde α_j es un término compuesto de μ_j y β_1 y el “-” se reemplaza por un “+” esta fórmula no afecta la estimación de la probabilidad máxima.

$$\log \left[\frac{\text{Prob}(y \leq j|x)}{1 - \text{Prob}(y \leq j|x)} \right] = \alpha_j + \beta \cdot x_i$$

Basado en el modelo de respuesta general ordinal de j categorías para la k variables explicativas y $i = 1, 2, \dots, n$ individuos el modelo es:

$$\text{Prob}(y \leq j|x) = \alpha_j + \beta \cdot x_i$$

Donde β es un vector de $k \times 1$, del intercepto de los parámetros, x_i es un vector de $1 \times k$ de las variables del i -ésimo individuo, α_j es el intercepto del j -ésimo valor categórico. Para tres categorías la probabilidad de la i -ésima es

$$\text{Prob}(y_j = 1|x_i) = F(\alpha_1 + \beta \cdot x_i)$$

$$\text{Prob}(y_j = 2|x_i) = F(\alpha_2 + \beta \cdot x_i)$$

$$\text{Prob}(y_j = 3|x_i) = 1 - F(\alpha_2 + \beta \cdot x_i)$$

Donde F es la función acumulativa Logit,

$$F(x) = \frac{1}{1 + e^{-(\alpha_j + \sum \beta \cdot x_i)}}$$

Ya que la probabilidad de transición es el objeto de este estudio, se considera más apropiado trabajar con los betas, en vez de usar las proporciones de probabilidades; es decir, se usarán los betas y no los odd ratios del modelo, lo cual es una práctica común de los modelos Logit (DeMaris 1993; Liao 1994).