



Una aplicación del modelo Lee-Carter para la modelación de la dinámica de la deserción estudiantil universitaria

Juliana Pavas Herrera
Jessica Andrea Vergara Hurtado

Universidad del Valle
Facultad de Ingeniería, Escuela de Estadística
Santiago de Cali, Colombia
2022

Una aplicación del modelo Lee-Carter para la modelación de la dinámica de la deserción estudiantil universitaria

Juliana Pavas Herrera
Jessica Andrea Vergara Hurtado

Trabajo de grado presentado como requisito parcial para optar al título de:
Estadístico

Director:
Ph.D. Jaime Mosquera Restrepo
Codirector:
M.Sc. Gabriel Conde Arango

Universidad del Valle
Facultad de Ingeniería, Escuela de Estadística
Santiago de Cali, Colombia
2022

Agradecimientos

En primera medida dar gracias a mis familiares por el apoyo incondicional y la paciencia que me han brindaron a lo largo de mi formación profesional, seguido agradecer de forma muy especial al director de tesis el profesor Jaime Mosquera, por su acompañamiento y contribución al presente trabajo, además, agradecer a Guillermo Arango, a la Vicerrectoría Académica de la Universidad del Valle quien suministró los datos y el apoyo en el entendimiento. Tercero a todos mis compañeros de la universidad que me acompañaron en el proceso de formación académica y dejaron enseñanzas en mi vida. Por último, a todos los profesores de la Escuela de Estadística.

Este Trabajo de Grado me gustaría dedicarlo a la memoria de mi compañera Danna Valentina Gonzáles Marín.

Juliana Pavas Herrera

Agradecimientos

Gracias a Dios por permitirme llegar a este punto y llenarme de nuevas oportunidades y bendiciones.

A mis padres y hermana por su apoyo incondicional, por creer en mi en todo momento, este logro va dedicado a ellos principalmente, al profesor Jaime Mosquera por su paciencia y acompañamiento en todo este proceso, agradecer a Guillermo Arango por el suministro de la información y la colaboración para el entendimiento de los mismos, a los profesores de la escuela de estadística los cuales contribuyeron a mi crecimiento profesional como personal y por último a todos mis compañeros de universidad que dejaron huella en mí, en especial a Danna Valentina Gonzalez Marin *Q.P.D* por dejarme tantos bonitos recuerdos y tener el placer de haber recorrido gran parte de este camino a su lado.

Jessica Andrea Vergara Hurtado

Resumen

Este trabajo tiene como propósito caracterizar y analizar la estructura y tendencia temporal de la deserción estudiantil en el programa académico de Estadística de la Facultad de Ingeniería de la Universidad del Valle durante los periodos 2000-2019 a partir del modelo Lee Carter, usado principalmente en descripción, pronóstico de mortalidad y esperanza de vida, este modelo se ajusta a partir de tablas de mortalidad académica las cuales se construyeron a partir de la información suministrada por registro académico, finalmente se realiza una comparación y análisis de resultados por género.

Palabras clave: *Tablas de mortalidad, modelo Lee-Carter, modelo lineal no generalizado, tendencia temporal, modelo de regresión, modelo lineal generalizado.*

Abstract

The purpose of this work is to characterize and analyze the structure and temporal trend of student desertion in the academic program of Statistics of the Faculty of Engineering of the Universidad del Valle during the periods 2000-2019 based on the Lee Carter model, used mainly in description and prognosis of mortality and life expectancy, this model is adjusted from academic mortality tables which were built from the information provided by the academic record, finally a comparison and analysis of results by gender is carried out.

Keywords: *housing supply, external covariates, web scraping, regression model, generalized linear model.*

Contenido

1. Introducción	2
1.1. Planteamiento del Problema	3
1.2. Justificación	4
1.3. Objetivos	6
1.3.1. Objetivo General	6
1.3.2. Objetivos Específicos	6
2. Antecedentes	7
3. Marco Teórico	13
3.1. Marco Conceptual	13
3.2. Marco Teórico Estadístico	15
3.2.1. Tablas de mortalidad	15
3.2.2. Modelo Lee-Carter	17
3.2.3. Modelos Lineales Generalizados	19
4. Metodología	23
4.1. Población de estudio	23
4.2. Construcción y depuración de base de datos de deserción	25
4.3. Cálculo y construcción de tablas de mortalidad abreviadas	27
4.4. Análisis exploratorio de datos	29
4.5. Ajuste del Modelo Lee-Carter	29
5. Resultados	32
5.1. Análisis Exploratorio	32
5.1.1. Análisis Exploratorio Mujeres	34
5.1.2. Análisis Exploratorio Hombres	35
5.2. Ajuste del Modelo Lee Carter	37
5.2.1. Ajuste del Modelo Lee Carter- Mujeres	38
5.2.2. Ajuste del Modelo Lee Carter- Hombres	40
5.2.3. Ajuste del Modelo Lee Carter- Contraste Hombres vs Mujeres	41
5.3. Análisis de Residuos	42

6. Conclusiones y recomendaciones	44
6.1. Conclusiones	44
6.2. Recomendaciones	45
Bibliografía	46
A. Anexo: Estructura de base de datos de la Oficina de Registro Académico de la Universidad del Valle	49

Lista de Figuras

2-1. Clasificación de la deserción de acuerdo al tiempo (Castaño et al., 2008) . . .	8
4-1. Flujograma de metodología	23
4-2. Ejemplo de imputación de datos	28
4-3. Tabla de mortalidad dinámica abreviada para la mortalidad académica en el pregrado Estadística de la Universidad del Valle (Periodo)	29
5-1. Porcentaje de deserción de estudiantes del programa académico Estadística de la Universidad del Valle	32
5-2. Porcentaje de deserción de estudiantes pregrado estadística Universidad del Valle por periodo y edad	33
5-3. Porcentaje de deserción de estudiantes mujeres del programa académico Estadística de la Universidad del Valle	34
5-4. Porcentaje de deserción de las estudiantes mujeres del programa académico Estadística de la Universidad del Valle por periodo y edad	35
5-5. Porcentaje de deserción de estudiantes hombres del programa académico Estadística de la Universidad del Valle	36
5-6. Porcentaje de deserción de las estudiantes hombres del programa académico Estadística de la Universidad del Valle por periodo y edad	36
5-7. Estimaciones de parámetros a_x, b_x y k_t del modelo Lee Carter. Datos generales	38
5-8. Estimaciones de parámetros a_x, b_x y k_t del modelo Lee Carter. Mujeres . . .	39
5-9. Estimaciones de parámetros a_x, b_x y k_t del modelo Lee Carter. Hombres . . .	40
5-10. Estimaciones de parámetros a_x, b_x y k_t del modelo Lee Carter. Contraste Hombres Vs Mujeres	41
5-11. Residuos del modelo Lee-Carter por género, periodo y edad	42
A-1. Estructura de datos de Registro Académico de la Universidad del Valle de código de estudiantes y periodos matriculados	49
A-2. Estructura de datos de Registro Académico de la Universidad del Valle de código de estudiantes y fecha de grado	50
A-3. Estructura de datos de Registro Académico de la Universidad del Valle de código de estudiantes, sexo, fecha de nacimiento y estrato	50
A-4. Estructura de datos con categorías para cada estudiante	51

Lista de Tablas

3-1. Bajos rendimientos respecto al número de asignaturas matriculadas (Consejo, 1997)	14
3-2. Estructura general de una tabla de mortalidad (Ortiz et al., 2012)	15
3-3. Estructura general de la tabla de mortalidad abreviada (Ortiz et al., 2012) .	16
3-4. Estructura general de tabla de mortalidad dinámica (Ortiz et al., 2012) . . .	16
3-5. Estructura MLG (Cayuela, 2009)	19
3-6. Funciones de enlace comúnmente utilizadas en modelos lineales generalizados (Cayuela, 2009)	20
4-1. Agrupación de semestres para la construcción de las tablas de mortalidad abreviadas.	26

1. Introducción

El término “deserción” en la educación superior, hace referencia al abandono prematuro de un programa de estudios provocado por factores que se generan tanto al interior del sistema educativo, como también, por relaciones externas con la sociedad, la familia y el entorno. De acuerdo con la educadora Himmel (2002), *"la deserción universitaria se ha convertido en un problema estructural de la educación superior, el cual afecta directamente al territorio nacional, lo que impacta en el desarrollo económico, social y científico del país"*. Hasta este momento, una de las dificultades para estudiar la deserción, radica en que cada Institución de Educación Superior (IES) maneja los datos con metodologías propias y con estructuras particulares, que en muchas ocasiones no permiten hacer el seguimiento individual a cada estudiante matriculado, con el fin de detectar los factores que ponen en riesgo la deserción estudiantil, y, encontrar las señales de alerta a tiempo para asegurar su permanencia en las aulas.

Para superar la dificultad anterior, fue diseñado el Sistema para la Prevención y Alerta de la Deserción en Instituciones de Educación Superior (SPADIES), como una herramienta informática diseñada para hacer seguimiento a las cifras de deserción de estudiantes de la educación superior en Colombia. Este sistema utiliza como insumo los datos suministrados por las IES, que han sido previamente estandarizados y ordenados, para evaluar el efecto de algunas variables determinantes de la deserción, estudiar las causas y evaluar el riesgo. Los datos almacenados en el SPADIES pueden ser consultados por cada IES, quienes podrán utilizarlos como insumos para analizar el fenómeno de deserción en el contexto particular de cada Institución.

La deserción estudiantil es un fenómeno antiguo, que ha sido objeto de investigación a nivel mundial durante un largo tiempo, dando lugar a distintos tipos de modelos conceptuales, que han tratado de explicar y describir este fenómeno. En la literatura, se encuentran autores que se centran en modelos basados en una perspectiva económica. Por ejemplo, Tinto (1975) se basa en un modelo de desgaste que atribuye mayor importancia relativa a factores externos a la Universidad; Bean (1985) construye un modelo del síndrome de deserción que se basa en la forma cómo el estudiante asume diferentes aspectos académicos, sociales y personales. Adicionalmente, se encuentran trabajos nacionales como Escobar et al. (2006) y Simonds (2012) quienes han tratado la deserción a partir de modelos estadísticos no paramétricos y de riesgos de supervivencia. usados para estudiar el fenómeno de la deserción en los programas académicos en la Universidad del Valle.

Más recientemente, Ochoa (2015), consideró el ajuste de un modelo de Lee-carter para el estudio de la deserción universitaria en la Universidad Nacional, sede Medellín, Colombia. El modelo de Lee-carter, fue propuesto en el año 1992 como una alternativa para estudiar los cambios temporales en las tablas de mortalidad de la población. La propuesta de Ochoa es novedosa, en el sentido de que el ajuste de un modelo LC para la deserción, permite identificar instantes del tiempo en los cuales la estructura de la deserción ha presentado cambios significativos. Al tiempo, que permite identificar los puntos críticos donde se presenta mayor incidencia de la deserción en la estructura curricular de un programa académico o de una institución. Motivado por lo anterior, en el presente trabajo se estudia el ajuste del modelo de Lee-Carter, para el estudio de la deserción universitaria en la Universidad del Valle en el programa académico de Estadística, debido a la cercanía de las autoras con el pregrado.

En este documento se construyen las tablas de mortalidad de deserción académica donde serán analizadas a partir del modelo Lee-Carter ya que en la actualidad los modelos paramétricos suelen ser más utilizados pues es habitual que los parámetros permiten realizar predicciones sobre la mortalidad futura de una manera más sencilla ((Díaz Rojo and Debón Aucejo, 2016)). A través del modelo LC se podrá determinar la estructura de la deserción en el programa académico de Estadística y su tendencia temporal. Además, permitirá identificar los semestres donde se encuentra principalmente acumulado el fenómeno de deserción. puesto que este estudio podría llegar a identificar puntos críticos de deserción que contribuiría a que el programa académico pueda intervenir o evaluar los cambios que han ocurrido en las reformas curriculares para plantear opciones de mejoramiento.

La estructura del documento se divide en 6 Capítulos, en el Capítulo 1 se realiza una breve introducción a la problemática de interés, explicando las razones e importancia de este estudio, en el Capítulo 2, se presenta una recopilación de estudios relacionados con la deserción académica y el modelo Lee-Carter, en el Capítulo 3, se exponen las herramientas estadísticas y aspectos teóricos soporte para el desarrollo de este proyecto de investigación, en el Capítulo 4 se describe metodología implementada, en el Capítulo 5, se describen los resultados obtenidos, finalmente en el Capítulo 6 se presentan las conclusiones finales y recomendaciones para estudios posteriores.

1.1. Planteamiento del Problema

La deserción estudiantil en general está presente en todos los niveles educativos y afecta al sistema en costos económicos y sociales, puesto que implica incumplimiento de los objetivos institucionales respecto a la formación y, además, se ve comprometida la calidad y los índices de eficiencia académica. Es decir, no es un concepto exclusivo de la educación superior.

Uno de los principales problemas que enfrenta el sistema de educación superior colombiano corresponde a los altos niveles de deserción académica en el pregrado. En el año 2008, el Ministerio de Educación Nacional (MEN) (Ruiz et al., 2009), indicó que la mayoría de casos de deserción universitaria en Colombia se deben a que las bases en la educación secundaria son débiles, y, que los procesos de admisión en las universidades fallan constantemente en detectar a quienes están preparados realmente para acceder a los estudios de educación.

Según el Ministerio de Educación de Colombia, para el año 2016, el indicador de deserción estructural a nivel universitario fue de 45.1 % (Educación, 2016). De acuerdo al indicador se muestra que cerca de la mitad de los estudiantes que ingresan a la educación superior no culminan este proceso de manera exitosa. Conforme con las cifras del Banco Mundial, para el año 2018, Colombia se ubicó como el segundo país de América Latina con mayor tasa de deserción en educación superior.

En el caso de la Universidad del Valle, Escobar et al. (2006) indicaron que la deserción entre los años 1994 al 2002 en la Facultad de Ingenierías la tasa de deserción supera el 50 %. Para el año 2010, Oficina de Planeación (2011) determinó que la tasa promedio de deserción en la Universidad del Valle se encontraba aproximadamente en el 42 %. Particularmente, en el Programa de pregrado en Estadística, Simonds (2012), encontraron que los índices de deserción en han oscilado entre el 35 % y 57 %, una cifra superior al promedio nacional y de la Universidad del Valle, en el cual los primeros cuatro semestres se presentan un 69,9 % de las deserciones definitivas.

De ahí la importancia del presente trabajo, pues es indispensable buscar soluciones que puedan prevenir la deserción superior que a corto plazo afecta el desarrollo socioeconómico, no solo de las universidades, sino también del país.

Finalmente, la deserción estudiantil en la Universidad del Valle ha sido investigada a partir de varios enfoques y distintos modelos. Sin embargo, hasta el momento no existe ningún estudio que incluya el modelo Lee-Carter, el cual permite identificar instantes del tiempo en los cuales la estructura de la deserción ha presentado cambios significativos, al tiempo que permite identificar los puntos críticos donde se presenta mayor incidencia de la deserción en la estructura curricular de un programa académico o de una institución.

1.2. Justificación

De acuerdo con estadísticas del Ministerio de Educación Nacional del año 2009 (Ruiz et al., 2009), de cada cien estudiantes que ingresan a una Institución de Educación Superior (IES), cerca de la mitad abandona sus estudios y no logra culminar su ciclo académico para obtener la graduación. En este sentido, las pérdidas financieras y sociales generadas por los estudiantes desertores son altas para la sociedad, las IES y para las familias de los estudiantes implicados.

A pesar de que el problema de la deserción ha sido estudiado por muchos autores, las causas de la misma no son universales, por el contrario, corresponden a situaciones muy particulares del entorno de la población atendida por cada IES. De manera que no ha sido posible identificar una causa determinante de la deserción para cualquier IES. En consecuencia, se sugiere que cada institución, incluso cada programa académico, estudie de forma particular el fenómeno para la población atendida, considerando inicialmente múltiples factores que involucren aspectos académicos, sociales, económicos, temporales y de estructura curricular.

Anteriormente, Ochoa (2015) estudió la dinámica de la deserción, a través del ajuste del modelo de Lee-Carter. Este modelo permitió identificar los instantes de tiempo y los semestres académicos con mayor impacto de deserción para la Institución estudiada. Específicamente, realiza un estudio para la población estudiantil de pregrado en la Universidad Nacional de Colombia sede Medellín a través de 34 cohortes, exactamente desde 02/1989 hasta 01/2006 usando el modelo de Lee-Carter. Usando este modelo concluye que la evolución de la población universitaria refleja un incremento en la deserción en los primeros semestres a partir de 2001, y al mismo tiempo, una disminución del tiempo requerido para graduarse en 10 semestres.

Motivados por los hallazgos de Ochoa (2015), en el presente estudio se hace uso del modelo Lee-Carter para llevar a cabo la investigación, ya que su estructura involucra parámetros que permiten analizar el comportamiento temporal de la deserción, como también el comportamiento de la deserción en el perfil del avance del estudiante a través de los semestres que conforman el currículum académico.

Particularmente, siguiendo las pautas de Ochoa (2015), en el presente trabajo se estudia la dinámica de la deserción para el programa académico de Estadística de la Universidad del Valle. Para realizar la modelación se explora el uso de dos fuentes de datos de deserción: EL SPADIES y los datos de la oficina de Registro Académico de la Universidad del Valle. Desde ambas fuentes se incluyen los datos de los estudiantes matriculados en el programa en el periodo 2000-1 hasta 2019-2.

A pesar de que se desarrolla la implementación para un caso particular, se espera que la metodología desarrollada sea replicable para otros programas académicos de la misma institución, incluso de otras instituciones, y contribuir con ello a identificar las causas de la deserción y su mitigación. La razón principal del planteamiento expuesto, es el escaso conocimiento que se tiene en el país sobre los ciclos de la deserción. Para ello, se ha decidido estudiarla desde una perspectiva de su evolución temporal, con el fin de encontrar nuevas alternativas para mitigar y, al mismo tiempo, implementar la visión del análisis temporal, la cual puede agregar utilidad a los datos arrojados por el sistema de información SPADIES.

1.3. Objetivos

1.3.1. Objetivo General

Caracterizar, mediante el ajuste del modelo de Lee-Carter, la estructura de la deserción en el programa académico de Estadística y su tendencia temporal.

1.3.2. Objetivos Específicos

- Construir las tablas dinámicas de mortalidad asociadas al fenómeno de la deserción estudiantil para el programa académico de Estadística.
- Caracterizar el comportamiento estructural y la tendencia temporal de la deserción en el programa académico de estadística, mediante un análisis exploratorio de datos.
- Modelar, a través del modelo de Lee Carter, el comportamiento estructural y la tendencia de la deserción en el programa académico de Estadística.

2. Antecedentes

En esta sección se presenta una revisión bibliográfica del estado del arte, que constituye varios desarrollos de investigación acerca de la deserción estudiantil universitaria a nivel nacional y particularmente en la Universidad del Valle. Con el fin de indagar acerca de las metodologías utilizadas en trabajos previos para el estudio del fenómeno de la mortalidad académica. Es importante resaltar que este estudio se realizará por medio del ajuste del Modelo Lee-Carter por lo que se exponen estudios que aportan significativamente a la realización de este trabajo de grado.

Castaño et al. (2008)

En este trabajo se realizó un estudio en la Facultad de ciencias económicas y la Facultad de Ingeniería de la universidad de Antioquia, con información de la segunda cohorte de 1996. Las fuentes de información empleadas fueron el Sistema de Información Matrícula y Registro, el Módulo de Inscripción y Selección Sistemática y una encuesta que se realizó para obtener datos no incluidos en los sistemas de información de la universidad, con el fin de identificar los factores influyentes en la deserción universitaria.

Se presenta el análisis y definición de los determinantes de la deserción estudiantil, realizado a partir de la aplicación de los modelos de riesgo proporcional en tiempo discreto. Esta metodología aplicada al problema del tiempo de deserción, permitió hacer un seguimiento de los estudiantes desde el inicio de sus estudios hasta que se presenta el evento de deserción.

Esta investigación define el abandono estudiantil en función del tiempo de la siguiente manera:

- **Deserción Precoz:** El individuo que, habiendo sido aceptado por la universidad, no se matricula.
- **Deserción Temprana:** Aquel que abandona sus estudios en los primeros semestres de la carrera.
- **Deserción Tardía:** Quien abandona los estudios en los últimos semestres, es decir, una vez cursados al menos la mitad de los semestres establecidos en el programa académico (Figura 2-1).

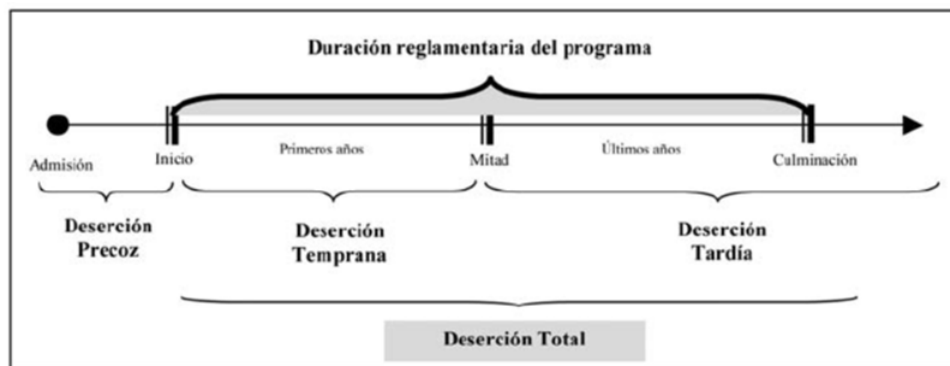


Figura 2-1.: Clasificación de la deserción de acuerdo al tiempo (Castaño et al., 2008)

Finalmente, una de sus conclusiones indica que la decisión de los alumnos que desertan no sólo está determinada por la ausencia de interés o capacidad ante las exigencias del trabajo académico, sino también, por un amplio conjunto de factores extra académicos. Adicionalmente, es importante que las instituciones de educación superior, en particular las públicas, emprendan campañas pedagógicas que recalquen la importancia de terminar los estudios en el tiempo teóricamente establecido.

Escobar et al. (2006)

Determinaron como objetivo mostrar el comportamiento de la deserción, permanencia y graduación para las cohortes que ingresaron en 1994 hasta 2001, cerrando la ventana de observación en el año 2006. La población de estudio fue seleccionada en los semestres comprendidos entre 1994-II y 2006-I, con la información de los estudiantes matriculados en la Universidad del Valle en la sede de Cali de todos los programas académicos.

En esta investigación definen la deserción como una situación a la que se enfrenta un estudiante cuando aspira y no logra su proyecto educativo, considerando como desertor a aquel estudiante de un programa académico que abandona definitivamente sus estudios.

Los factores que se asocian a la deserción teniendo en cuenta los resultados de la estimación del modelo de duración de riesgo proporcional en tiempo discreto de Prentice and Gloeckler (1978). Sugiere que los estudiantes de sexo masculino, solteros, que viven independientemente y que sufrieron algún tipo de calamidad doméstica, al parecer, tienen un mayor riesgo de deserción. Adicionalmente, la edad de los alumnos parece incidir en el riesgo de desertar, y éste decrece marginalmente al aumentar la misma.

Para el análisis del modelo de duración, estimaron las funciones de supervivencia de los estudiantes mediante el método no paramétrico de Kaplan-Meier y el riesgo de supervivencia

a partir de los modelos paramétricos de duración siguiendo distribuciones Weibull para la función de riesgo base. En el ajuste del modelo se incluyeron variables socioeconómicas, relacionadas con el hogar y la procedencia de los estudiantes, además de variables académicas e institucionales.

Mosquera Daza and Valdés Velásquez (2008)

En el trabajo de grado titulado “Estudio de permanencia de los estudiantes de los programas tecnológicos de la facultad de Ingeniería de la Universidad del Valle, Sede Meléndez, para el periodo 2002-2006” donde el propósito principal es identificar la medida del riesgo de retiro y los factores asociados al retiro de los estudiantes de los Programas Tecnológicos de la Facultad de Ingeniería de la Universidad del Valle, sede Meléndez, el estudio se realizó a partir de la base de datos obtenida a través de la Oficina de Registro Académico de la Universidad del Valle. Esta base resume la trayectoria de los estudiantes de las tecnologías para las cohortes entre agosto-diciembre de 2002, enero-mayo de 2003 y agosto-diciembre de 2006.

Para el desarrollo se utiliza la metodología del Análisis de Supervivencia para estimar el riesgo de retiro y se definen dos eventos de interés los cuales son: deserción y graduación. Para cada evento se obtiene la función de supervivencia por medio del estimador no-paramétrico propuesto por Kaplan y Meier, también se utiliza el modelo semiparamétrico de Cox para determinar las variables significativas dentro del modelo que permitan explicar el retiro de los estudiantes de los Programas Tecnológicos.

Como resultado obtuvieron que las tasas de deserción son muy altas (57.5 %) donde el promedio de deserción de los Programas Tecnológicos al primer semestre es del 25.2 %; por otra parte las tasas de graduación muy bajas (21.6 %).

Respecto al evento asociado al retiro (graduación), desarrollado por la función de supervivencia por medio del estimador no-paramétrico propuesto por Kaplan y Meier. Se tiene que la mayor ocurrencia del evento retiro (graduación) se presentó en el octavo semestre de permanencia en el Programa Tecnológico pues la proporción de estudiantes que logró sobrevivir en este tiempo es del 38.2 %.

Simonds (2012)

En este trabajo de grado titulado “Análisis de la deserción de los estudiantes del programa de estadística de la Universidad del Valle 1999 - 2011” tiene como objetivo la construcción de una herramienta metodológica que permita clasificar a los estudiantes del Programa Académico

de Estadística de la Universidad del Valle, según el nivel de riesgo de deserción en el programa de durante el periodo 1999-2011, además se permite identificar los factores asociados a la deserción definitiva del programa. Para el proceso de análisis de datos y construcción de los mismos, se basaron en la información suministrada por Registro académico, esta base de datos contiene factores académicos, individuales y socioeconómicos de los estudiantes al ingreso al programa.

Se realiza un perfil de cada estudiante al ser admitido al programa académico, de los desertores parciales, definitivos y se realiza una comparación entre ellos. Mediante el estimador de curvas de supervivencia de Kaplan-Meyer se identifican posibles factores asociados al evento deserción definitiva, se ajusta un modelo de riesgos proporcionales para identificar los factores que tienen mayor incidencia en que ocurra el evento y se validan los supuestos del modelo.

Como resultados al ajustar el modelo de riesgos proporcionales, concluye que las variables: Género, Reside en Cali, Tiempo de espera y Edad (Edad centrada en 18 años) son las más importantes al explicar el evento deserción definitiva de los estudiantes del programa académico de estadística en el periodo 1999-2011. Es de resaltar que el riesgo de desertar de los hombres es 1.7 veces mayor que el de las mujeres. Al clasificar riesgo de desertar definitivamente según las condiciones de ingreso de un estudiante, se tiene que a mayor tiempo de espera desde que el estudiante se gradúa del colegio e ingresa a la universidad, aumenta el riesgo de desertar definitivamente, principalmente después de los 5 años, probabilidad que se hace mayor si el estudiante que ingresa es hombre, no reside en Cali y su edad es mayor a 18 años, se podría asociar con un estudiante entre más joven tiene mayor posibilidad de mantenerse en un sistema educativo que exige dedicación de tiempo completo. Los niveles de riesgo más bajos están asociados a estudiantes que son mujeres, residen en Cali, tienen un tiempo de espera entre 0 y 1 año y su edad no supera los 26 años.

Ochoa (2015)

En este estudio se presentan dos aplicaciones del modelo Lee-Carter, en primera instancia se estima un modelo para la dinámica de mortalidad en Colombia para los años 1951-1999, cuyo objetivo es analizar las diferentes metodologías y herramientas de estimación del modelo Lee-Carter aplicado a datos de mortalidad poblacional en datos demográficos colombianos. Acto seguido se analiza la dinámica de la mortalidad académica, modalidad pregrado en la Universidad Nacional de Colombia sede Medellín en los periodos semestrales entre el 2002 y 2006-2007. Para la modelación de la dinámica de mortalidad, se ilustra que el proceso de estimación de los parámetros se determina por medio de la metodología de la descomposición singular y de modelos lineales generalizados.

La estimación implementada para el modelo se realiza de tres diferentes maneras a través del programa estadístico R:

1. SVD - Descomposición singular de matrices a través de la función `svd` de R o la librería `demography`.
2. GLM - Modelos lineales generalizados a través de la librería `ILC`.
3. GNM - Modelos no lineales generalizados a través de la librería `GNM`.

En los parámetros del modelo bajo las tres formas, se obtienen estimaciones muy similares en las librerías `Demography` e `ILC`, con alguna diferencia relevante para los datos pertenecientes a las edades mayores. Por otro lado, la librería `GNM` tiende a desviarse hacia una menor fuerza de mortalidad media $\text{Ln}(\alpha_x)$ en las edades entre 0-40 y a tener una mayor magnitud en las edades de 40-80 años, lo que en primera instancia nos indica que la librería `GNM` logra representar una mayor variabilidad en rangos extremos de edades tanto para hombre como para mujer.

Para ambos géneros, Hombres y Mujeres, se presentan los residuos tipo Deviance y tipo Pearson, bajo las metodologías planteadas librería `Demography` (que usa descomposición singular), `ILC` (máxima verosimilitud) y `GNM` (minimiza el estadístico Deviance). Se concluye que la mejor estimación es la realizada por el modelo log-bilineal de Poisson a través del uso de la herramienta `GNM`.

Respecto a la dinámica de la deserción, encuentran que la evolución de la población universitaria refleja un incremento en la deserción en los primeros semestres a partir de 2001, y al mismo tiempo, una disminución del tiempo requerido para graduarse en 10 semestres.

Díaz Rojo and Debón Aucejo (2016)

Este trabajo estudió el comportamiento de la mortalidad en Colombia para el período 1973-2005 utilizando tablas de mortalidad construidas a partir de *Latin American Human Mortality Database* y modelar la mortalidad en Colombia con el modelo de Lee-Carter.

Inicialmente, para la realización de la estimación del modelo Lee-Carter es indispensable la construcción de las tablas de vida conocidas como tablas de mortalidad, la cual es la herramienta comúnmente usada para medir las probabilidades de vida o de muerte de una población. Para la construcción de la tabla de vida es necesario contar con los datos de población provenientes de censos y de los registros de mortalidad existentes tales como cantidad de años cumplidos del individuo, número de sujetos muertos en ese rango de edad durante el año analizado, promedio de años vividos. Para la construcción de las tablas de vida, en este estudio se agruparon edades en quinquenios desde 0 hasta 84 años, con el resto de edades superiores a 84 agrupadas en una única categoría (85+).

Las autoras proponen tres metodologías de estimación de las tablas de mortalidad: el primero es el modelo Lee-Carter (LC) habitual propuesto por Lee y Carter (1992), que expresa la tasa central de mortalidad como una función exponencial que depende de la edad y del tiempo. El segundo, el modelo Lee-Carter con dos términos (LC2), el cual es un modelo alternativo que consiste en agregar un término a la expresión con el objetivo de mejorar su ajuste. Y por último, el Modelo Edad-Periodo-Cohorte (APC), que suele analizar las tendencias de la incidencia y de la mortalidad para valorar el efecto temporal de la ocurrencia de un evento.

Como resultado, encontraron que la mortalidad colombiana muestra similitudes con el comportamiento de la mortalidad para países desarrollados; Los gráficos de los residuos en función del tiempo, la edad y la cohorte muestran, en todos los casos, que los datos de mortalidad de las mujeres son los que presentan un mejor ajuste de los modelos, siendo el modelo LC2 el que ofrece mejor ajuste, puesto que presenta valores menores para ambas medidas tanto para hombres como para mujeres y es el que mejor explica los efectos de la mortalidad para ambos sexos en Colombia. Por lo anterior, sugiere utilizar el modelo de LC2 para futuras predicciones.

Uribe Gil (2017)

Realiza un estudio que expone la tendencia de mortalidad de Hombres y Mujeres, en Santiago de Cali, mediante la aplicación del Modelo Lee-Carter. El estudio pretende llevar a cabo el análisis del pronóstico de las tasas de mortalidad de Cali de 1985 a 2030 y la exploración de las tendencias de las tablas de vida. La metodología empleada evidencia reducciones de las tasas de mortalidad de los periodos estudiados, a excepción de los hombres de los grupos de edad de 14 y 30 años. Se destaca que una de las ventajas del modelo Lee-Carter es que permite el pronóstico de tasas que dependen de una tendencia estocástica subyacente.

3. Marco Teórico

En esta sección se presenta el marco teórico de la investigación. En la primera parte (Sección 3.1) se muestra el marco conceptual, en el cual se incluyen definiciones importantes acerca de la deserción académica. En la segunda parte (Sección 3.2), se encuentra el marco teórico estadístico, donde se presenta la teoría fundamental para el estudio y su respectivo desarrollo matemático del modelo Lee-Carter.

3.1. Marco Conceptual

Según Tinto (1989) ninguna definición puede captar en su totalidad la complejidad de este fenómeno que es la deserción académica, quedando en manos de los investigadores la elección de la definición que mejor se ajuste a sus objetivos y al problema a investigar. De esta forma, para la realización de esta investigación desde un punto de vista institucional, se define lo siguiente:

Desertor estudiantil:

Aquel estudiante que abandona la educación superior sin registrar grado (Castaño et al., 2008).

Existen dos tipos de deserción estudiantil en la Universidad del Valle las cuales son voluntaria e involuntaria, la primera según Escobar et al. (2006), definen que hay variables externas que determinan el riesgo de deserción de los estudiantes como el factor socioeconómico y otros menos explorados, tales como los factores individuales. Por último, respecto a la deserción involuntaria según el Consejo (1997) en el Capítulo VIII DE LOS BAJOS RENDIMIENTOS del Acuerdo 009 de Noviembre 13 de 1997 de la Universidad del Valle, define en los Artículos 59°, 62° y 63° lo siguiente:

Artículo 59°

Se consideran situaciones de bajo rendimiento académico las siguientes:

1. Para los estudiantes cobijados por el Acuerdo 001 o las normas que lo modifiquen, la pérdida del 50 % o más de los créditos matriculados.

2. Para los estudiantes no cobijados por el Acuerdo 001, la pérdida, bien sea por evaluación o por la no cancelación reglamentaria de un número de asignaturas igual o superior al fijado como en la Tabla **3-1** siguiente cuadro:

Tabla 3-1.: Bajos rendimientos respecto al número de asignaturas matriculadas (Consejo, 1997)

Número de asignaturas matriculadas	Número de asignaturas perdidas después de la habilitación
1	1
2	1
3	2
4	2
5	3
6	3
7	4

Para un número mayor de asignaturas se considerarán situación de bajo rendimiento académico la pérdida del 50 % o más de las asignaturas registradas.

3. La pérdida de una asignatura que se cursa en calidad de repitente por primera vez. Sin embargo, si el estudiante no ha estado en situación de bajo rendimiento su promedio acumulado es de tres punto cinco (3.5) o mayor, no habrá lugar a un bajo rendimiento.
4. La pérdida en un mismo período académico de dos (2) o más asignaturas que se repiten.
5. La pérdida de la habilitación de una asignatura que se cursa en calidad de repitente por segunda vez.

Artículo 62°

El estudiante que incurra en bajo rendimiento académico por segunda vez en un programa académico no podrá continuar en él, a menos que haya aprobado el 60 % de sus créditos o aprobado más del 60 % de las asignaturas, en el caso de estudiantes no cubiertos por el Acuerdo N.° 001 de 1993 del Consejo Superior y tengan un promedio acumulado no inferior a tres punto dos.

Artículo 63°

El estudiante que incurra en bajo rendimiento académico por tercera vez en un Programa Académico, no podrá continuar en él a menos que haya aprobado el 80 % de sus créditos,

o aprobado más del 80 % de las asignaturas, en el caso de estudiantes no cubiertos por el Acuerdo N.º 001 de 1993 del Consejo Superior o por las normas que lo modifiquen y tengan un promedio acumulado no inferior a tres punto cinco (3.5).

3.2. Marco Teórico Estadístico

3.2.1. Tablas de mortalidad

También llamadas tablas de vida, es una herramienta que usualmente se utiliza para medir las probabilidades de vida o de muerte de una población o un subgrupo en función de la edad para un período establecido (Debón et al., 2009). Una tabla de mortalidad puede interpretarse como un modelo que representa la distribución estadística del tiempo de sobrevivencia esperado de los miembros de un grupo determinado (Ortiz et al., 2012). En la Tabla **3-2** se observa la estructura de una tabla de mortalidad general donde x es la edad, t el tiempo y m_0 es la tasa de mortalidad en la edad x .

Tabla 3-2.: Estructura general de una tabla de mortalidad (Ortiz et al., 2012)

Edad	Tiempo
x	t
0	m_0
1	m_1
2	m_2
...	...

Un aporte importante de las tablas de mortalidad es que permiten construir indicadores que precisan con más exactitud los cambios e impactos de la mortalidad en determinados grupos (Arriaga, 1996).

Las tablas de mortalidad se clasifican según su estructura y aspecto de la siguiente manera:

Tabla de mortalidad completa

Cuando se construye presentando la información para cada una de las edades, es decir, con las edades año a año, desde el nacimiento hasta la última edad disponible (Díaz Rojo and Debón Aucejo, 2016).

Tabla de mortalidad abreviada

Cuando la información se presenta agrupada por rangos de edades, este tipo de tablas de vida se utilizan cuando se dispone de los datos de mortalidad en tasas referidas a grupos

quinquenales de edad y no las tasas de mortalidad de cada año de edad (Díaz Rojo and Debón Aucejo, 2016). En la Tabla **3-3** se encuentra la estructura de una tabla de mortalidad abreviada, siendo x la edad agrupada en quinquenios.

Tabla 3-3.: Estructura general de la tabla de mortalidad abreviada (Ortiz et al., 2012)

Edad	Tiempo
x	t
$[1 - 5]$	m_0
$[6 - 10]$	m_1
$[11 - 15]$	m_2
...	...

Tabla dinámica de mortalidad

Se asume que las tablas de mortalidad, además de depender de la edad, depende del año calendario en que se alcanza dicha edad, en un contexto dinámico se asume que la mortalidad es función tanto de la edad como del año calendario. Así, con el fin de reflejar de manera explícita la evolución de la mortalidad con el año calendario (Ortiz et al., 2012). En la Tabla **3-4** se encuentra la estructura de una tabla dinámica de mortalidad, siendo x la edad y t_i los tiempos correspondientes.

Tabla 3-4.: Estructura general de tabla de mortalidad dinámica (Ortiz et al., 2012)

Edad	Calendario				
x	t_0	t_1	t_2	t_3	...
0	m_{00}	m_{01}	m_{02}	m_{03}	...
1	m_{10}	m_{11}	m_{12}	m_{13}	...
2	m_{20}	m_{21}	m_{22}	m_{23}	...
...

Estructura de una tabla de mortalidad

Con base en la revisión de antecedentes relacionados con tablas de mortalidad, como la de Díaz Rojo and Debón Aucejo (2016) se obtienen estructura base:

- **Edad (x):** Define la cantidad de años cumplidos del individuo. Para tablas abreviadas se señala la edad inferior de cada uno de los intervalos de edad.
- **Amplitud del intervalo de edad:** Incluye los años en cada intervalo de edad. En las tablas completas $n = 1$.
- P_x : Población correspondiente al rango de edad en el año (calendario) analizado.

- d_x : Incluye el número de sujetos muertos en ese rango de edad durante el año analizado.
- nm_x : Tasa central de mortalidad, es el resultado entre el cociente de d_x sobre P_x .

$$nm_x = \frac{d_x}{P_x} \quad (3-1)$$

- na_x : Se refiere al tiempo promedio vivido con edad cumplida x por aquellos individuos de la cohorte ficticia que mueren con dicha edad.
- nq_x : Las probabilidades de muerte se calculan a partir de las tasas de mortalidad específicas para cada grupo etario. Debe interpretarse como la probabilidad de fallecer entre las edades del intervalo para el sujeto que ha sobrevivido hasta la edad x .

$$nq_x = \frac{n \cdot nm_x}{1 + (n - na_x) \cdot nm_x} \quad (3-2)$$

- nl_x : Representa el número de individuos de una cohorte ficticia inicial que llegan con vida a la edad x .
- nd_x : Número de defunciones de la cohorte ficticia inicial que tienen lugar en individuos entre las edades x y $x + n$.

$$nd_x = nl_x \cdot nq_x \quad nl_{x+n} = nl_x \cdot nd_x \quad (3-3)$$

3.2.2. Modelo Lee-Carter

Lee and Carter (1992) publicaron un nuevo método para obtener pronósticos a largo plazo del nivel y el patrón de edad de la mortalidad, basado en una combinación de métodos estadísticos de series de tiempo y un enfoque simple para tratar la distribución de la mortalidad por edades. El método describe el logaritmo de una serie temporal de tasas de mortalidad específicas por edad como la suma de un componente específico por edad que es independiente del tiempo y otro componente que es el producto de un parámetro variable en el tiempo, que refleja el nivel general de la mortalidad, y un componente específico por edad que representa la rapidez o la lentitud con la que varía la mortalidad en cada edad cuando cambia el nivel general de mortalidad.

En su propuesta de modelo, Lee-Carter expresan la tasa central de mortalidad (m_{xt}) como una función exponencial a través de la edad y el tiempo. La ecuación general es la siguiente:

$$m_{xt} = \exp(a_x + b_x k_t + \epsilon_{xt}) \quad (3-4)$$

de forma equivalente se aplica logaritmo a la ecuación inicial y se obtiene la siguiente forma

$$\ln(m_{xt}) = a_x + b_x k_t + \epsilon_{xt} \quad (3-5)$$

donde,

- m_{xt} = Es la variable a modelar, tasa de mortalidad para una edad y año determinado.
- a_x = Es el parámetro de forma dependiente de la edad.
- k_t = Es el índice que representa la tendencia de la mortalidad general a lo largo del tiempo.
- b_x = es el parámetro de sensibilidad que expresa cómo responde la tasa de mortalidad a cambios en k_t para un rango específico de edad.
- ϵ_{xt} = Representa el error aleatorio, con media 0 y varianza σ_x^2 , y representa el componente estocástico que no se explica por el modelo.

Este modelo debe tener ciertas restricciones propuestas por Lee and Carter (1992), debido al término bilineal $b_x k_t$ estos parámetros del modelo no serán identificables, es decir, no podrán ser estimados, ya que el modelo no varía respecto a transformaciones lineales. Lo cual se puede apreciar de la siguiente manera: Supongamos que los vectores a_x , b_x , k_t son una solución. Entonces para cualquier escala de c , $a_x - b_x c$, b_x , $k_t + c$ también deben ser una solución. Está claro que si a_x, b_x, k_t es una solución, entonces a_x , $b_x c$, k_t/c también lo es. Por lo tanto, k_t se determina como una transformación lineal, b_x se determina como una constante multiplicativa y a_x se determina como una constante aditiva. Por lo tanto, se deben normalizar los parámetros para así poder obtener un modelo identificable.

$$\sum_t k_t = 0 \quad \sum_x b_x = 1 \quad (3-6)$$

Como se observa en la Ecuación 3-4, este no tiene variables explicativas, por lo cual no se puede ajustar por medio de regresión habitual. Por lo tanto, no cumple con la forma de un modelo lineal general y se debe buscar alternativas que permitan realizar la estimación de los parámetros de interés. Actualmente, se presentan tres metodologías de estimación de los parámetros: 1. SVD (Descomposición singular) la cual está basado en la metodología planteada por Lee-Carter inicialmente, 2. GLM (Modelos lineales generalizados) donde la estimación se realiza a través de máxima verosimilitud y 3. GNM (Modelo lineales no generalizados) donde implementa el método de máxima verosimilitud para la estimación en dos pasos; primero usa Newton-Raphson de manera iterativa para cada parámetro, el segundo paso es utilizar Newton-Raphson para actualizar los parámetros a través de generalización de matrices inversas.

Los modelos no lineales generalizados (GNM) se pueden considerar una extensión de los modelos lineales generalizados, en la cual algunos términos del predictor son no lineales en los parámetros. Ocurre cuando la especificación matemática de μ_x no es lineal o linealizable, por lo tanto, se denomina como un modelo no lineal generalizado, y se hace necesario aplicar técnicas aptas para predictores no lineales (Morales, 2009).

3.2.3. Modelos Lineales Generalizados

Los modelos lineales generalizados (MLG) extienden las aplicaciones del modelo lineal general para poder manejar otro tipo de dato, de manera que la variable Y sigue alguna distribución de la familia exponencial con media μ , la cual es generalmente una combinación no lineal de los predictores. Estos predictores afectan la distribución de Y mediante una combinación lineal de las variables predictoras. En la Tabla 3-5 se representa la estructura del MLG.

Tabla 3-5.: Estructura MLG (Cayuela, 2009)

Modelo Lineal Generalizado (MLG)
$y_i = \sum_j B_j X_{ij} + \varepsilon_i$ $\mu_i = E(Y_i)$ $n_i = \sum_j B_j X_{ij}$ $n_i = g(\mu_i)$

Siendo:

- y_i : Vector de la variable respuesta.
- X_{ij} : Matriz de variables predictoras y covariables.
- B_{ij} : Vector de parámetros
- $g(\mu)_i$: Función de enlace.

Un modelo lineal generalizado tiene tres componentes (McCullagh and Nelder, 1989):

- **Componente Aleatoria:** Identifica la distribución de probabilidad de Y y la distribución de los errores.
- **Componente Sistemática:** Corresponde a las variables del modelo y su combinación lineal, el cual genera el predictor lineal.
- **Función de enlace:** Especifica la función que liga a los componentes aleatorios y sistemáticos. Indica la forma en que la variable respuesta se relaciona con el componente lineal de predictores.

Por lo anterior, la función de enlace transforma el valor medio, similar a como se realiza en una regresión cuando se presentan problemas de linealidad, pero tomando los valores estimados por el modelo mediante la transformación inversa de la función de enlace (Cayuela, 2009). Por lo tanto, el conjunto de variables aleatorias independientes $Y_1; Y_2; Y_3; \dots; Y_n$ proviene de distribuciones relacionadas con la familia exponencial, esta variación depende del valor que toma la variable de respuesta en la que cada distribución sugiere una función de enlace diferente, algunas de ellas se muestran a continuación en la Tabla **3-6**.

Tabla 3-6.: Funciones de enlace comúnmente utilizadas en modelos lineales generalizados (Cayuela, 2009)

Nombre	Función de enlace: $\eta = g(\mu)$	$\mu = g^{-1}(\eta)$
Identidad	μ	η
Logaritmica	$\log(\mu)$	$\exp(\eta)$
Logit	$\log(\frac{\mu}{1-\mu})$	$\frac{\exp(\eta)}{1+\exp(\eta)}$
Log-Log	$\log[-\log(\mu)]$	$\exp[-\exp(\eta)]$

Estos modelos GLM consideran algunos supuestos, tales como:

- Los datos de la variable respuesta son independientes.
- La variable dependiente no necesariamente sigue distribución normal, pero sí se distribuye como miembro de la familia exponencial.
- Los GLM no asumen una relación lineal entre la variable respuesta y las variables independientes, pero sí asume una relación lineal cuando el predictor se transforma por la función enlace.
- No se necesita satisfacer la homogeneidad de varianza, ni la normalidad de los errores (a menos que la distribución lo amerite).
- Se estima los parámetros por un método de máxima verosimilitud en lugar de mínimos cuadrados ordinarios (el modelo converge).

Evaluación de un Modelo Lineal Generalizado

Según Cayuela (2009) no existe un único modelo correcto, que hace referencia a que existen diferentes funciones de enlace que permiten estimar un modelo (mirar Tabla **3-6**). Quiere decir que, se debe identificar en primera instancia un modelo que explique la mayor proporción de la varianza, con base a que los parámetros del modelo deben ser significativos. Para la evaluación y ajuste de un MLG se tiene lo siguiente:

- **Ajuste del modelo:** Se debe tener en cuenta el test de significancia para los estimadores del modelo y la cantidad de varianza explicada. Esto en los MLG se conoce como devianza (deviance), es una medida de bondad de ajuste la cual nos da una idea de como es la variabilidad de los datos, para obtener una medida de la variabilidad explicada por el modelo, se debe de comparar la devianza del modelo nulo (Null deviance) con la devianza residual (Residual deviance), tal y como se observa en la Ecuación 3-7 siendo D2 es la devianza.

$$D^2 = \frac{\text{Devianza modelo nulo} - \text{Deviance residual}}{\text{Devianza modelo nulo}} \times 100 \quad (3-7)$$

- **Análisis de los residuos:** Permite evaluar la bondad de ajuste del modelo, y tener la certeza de que los resultados obtenidos puedan considerarse acertados, en la evaluación de los residuos o errores se puede detectar generalmente los siguientes puntos:
 - Si existe relación entre variable X y Y.
 - Si existe normalidad entre los errores.
 - Validar si existe homocedasticidad e independencia entre los errores.
- **Selección y comparación del mejor modelo:** Para esto, se escoge el mejor conjunto de variables predictoras del modelo por medio de 3 métodos:
 - **Método Forward (selección hacia adelante):** Inicialmente no se selecciona ninguna variable. Luego, se elige la variable más significativa dependiendo algún criterio estadístico. Seguidamente se selecciona la segunda variable más discriminante y así sucesivamente. Se finaliza cuando las variables que aún no son seleccionadas no son significativas.
 - **Método Backward (eliminación hacia atrás):** Inicialmente se seleccionan todas las variables. En cada paso se elimina la variable menos discriminante. El algoritmo finaliza cuando todas las variables que permanecen discriminan significativamente.
 - **Método Stepwise (o regresión por pasos):** Se presenta la combinación de los dos métodos anteriores. Se introduce o elimina una variable dependiendo de la significación de su capacidad discriminatoria. Una ventaja, es que permite corregir decisiones tomadas en pasos anteriores.

Finalmente, para la comparación y selección del modelo se utiliza el criterio de información de Akaike (AIC) y la razón de verosimilitud. Con la razón de verosimilitud se presenta la siguiente prueba de hipótesis:

$H_0 : M_0$ Es apropiado para explicar la variable de respuesta.

$H_1 : M_1$ Es el modelo mas apropiado para explicar la variable de respuesta.

Estadístico de prueba:

$$F_0 = \frac{D_{1M_0} - D_{1M_1}}{(gl_1 - gl_0)\hat{\phi}} \quad (3-8)$$

Donde D_{1M_0} corresponde a la devianza residual del modelo sin la variable a evaluar (M_0), D_{1M_1} denota la devianza residual del modelo más complejo (M_1); gl_0 y gl_1 corresponden a los grados de libertad de M_0 y M_1 respectivamente.

Posteriormente, para la decisión se calcula el *valor-p* = $P(F > F_0)$ con gl_0 y gl_1 , rechazando la hipótesis nula si valor $p < 0,05$, el cual fue el nivel de significancia considerado para dicha prueba.

Respecto al Criterio de Akaike mide el desajuste entre una distribución hipotética y una distribución teórica. Además, este sigue el principio de la parsimonia, es decir cuando el número de parámetros de un modelo k aumenta el AIC también. Por tanto, escoger el modelo que tiene el mínimo AIC supone elegir el modelo con el menor número de parámetros posible, cabe destacar que el AIC es una medida global de la bondad del ajuste de modelos estadísticos (Amaya Jiménez, 2018), en general el AIC se define de la siguiente manera:

$$AIC = 2k - 2\ln(L) \quad (3-9)$$

Donde k es el número de parámetros del modelo y $\ln(L)$ es la función de log-verosimilitud para el modelo estadístico.

4. Metodología

En este capítulo se presenta la metodología de la investigación. Inicialmente se denota las características de la población del estudio, seguidamente se conforma la base de datos de la investigación, detallando la técnica utilizada para la extracción de la información, la depuración y las variables de entorno obtenidas. Finalmente se detallan los procedimientos más importantes que se utilizarán para el análisis exploratorio y la modelación de la deserción de estudiantes discriminados en general, hombres y mujeres. En la Figura 4-1 se encuentra representado por medio de un flujograma el proceso que se realizó para en el trabajo de investigación por parte de las autoras.

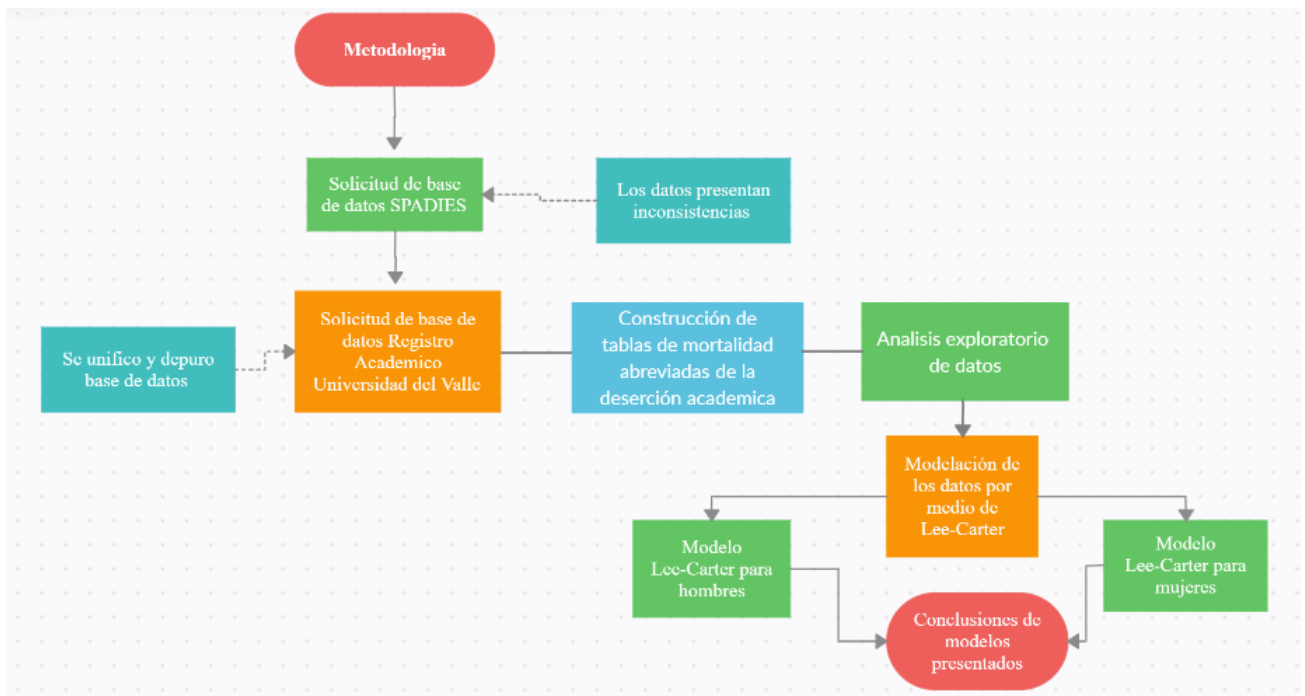


Figura 4-1.: Flujograma de metodología

4.1. Población de estudio

En esta investigación la población objeto de estudio serán los estudiantes de la Universidad del Valle de la Facultad de Ingeniería, del programa académico Estadística, matriculados

en los periodos de enero 2000 hasta diciembre 2019. Los datos inicialmente se solicitaron al sistema para la prevención de la deserción de la educación superior (SPADIES), sistema desarrollado por el Ministerio de Educación Nacional desde el 2002, que propone disminuir la deserción estudiantil en la educación superior como parte de la estrategia planteada para aumentar la cobertura, la calidad y la eficiencia educativa. En el proceso de validación correspondiente del SPADIES se concluye que la información suministrada presentó las siguientes inconsistencias :

- Muchos graduados del programa se encuentran en SPADIES como desertores del programa. También, algunos de ellos tienen un número de tarjeta de identidad como identificación y registran un número alto de periodos matriculados (≥ 9 semestres). Cuando se realiza el cambio de documento justo antes de graduarse (o que su grado se registró con cédula) este cambio no se reporta o no es leído por el SPADIES. En consecuencia, SPADIES no vuelve a encontrar el anterior número de identificación y se marca como desertor, sin detectar jamás su graduación. Esta situación se presenta con frecuencia y en consecuencia, infla la cifra de desertores en el programa académico.
- Respecto al ítem anterior, la gran mayoría de estudiantes presentan deserción en el periodo 2017-1. Se intuye que alguna novedad ocurrió en ese periodo académico con el proceso de notificación de cambios de documentos.
- Muchos estudiantes tienen presentan error en el registro del sexo en SPADIES. En consecuencia, la variable Género no informa correctamente, en la verificación se identificaron muchos casos registrados con el sexo contrario.
- Cuando un estudiante se gradúa, el SPADIES tiende a aumentar un periodo de matrícula adicional. Al parecer SPADIES toma la fecha de grado, que generalmente sucede un período académico posterior a la finalización del último período académico matriculado.

Por ejemplo, un estudiante, termina sus asignaturas en 2018-2, su grado fue en 2019-1, pero el estudiante no estuvo matriculado en ese periodo. El sistema le reporta el semestre 2019-1 como un periodo matriculado. Esto sucede para la gran mayoría de graduados, cuyos trámites de grado los extiende un semestre más. Esta situación genera un incremento sistemático en el tiempo de permanencia.

- Hay estudiantes que ingresan a un programa, se cambian a otro y luego se gradúan o desertan en el programa al que se trasladó. SPADIES los sigue hasta el grado o hasta la deserción en el otro programa, pero le suma todos los semestres matriculados al programa del primer ingreso. Esto incrementa falsamente el tiempo hasta la deserción en el programa de ingreso.

Por ejemplo: Un estudiante, es desertor en estadística en 6to semestre, pero se gradúa luego de 11 semestres en Administración. El dato de SPADIES indica que desertó de

estadística en el semestre 11. Al parecer no hay forma de identificar el instante exacto de tiempo de su deserción en un programa académico en específico

- Hay estudiantes activos que el sistema detecta como desertores y otros estudiantes que actualmente se encuentran activos el sistema solo les reconoce algunos periodos académicos.

Por ejemplo: Un estudiante, SPADIES marca 6 periodos matriculados y este actualmente acumula más de 12 semestres según Registro Académico .

- Los estudiantes admitidos por amnistía se encuentran en el registro como un nuevo ingreso, hacen dos matrículas y se gradúan, con lo cual el tiempo de graduación es falsamente dos. Lo mismo sucede con los estudiantes que hacen un nuevo ingreso después del año 2000, pero tienen historial previo antes de este año. El sistema solo le toma los semestres realizados después del nuevo ingreso.

En consecuencia, debido a la alta frecuencia de inconsistencias identificadas, se concluye que hay fallas estructurales en la información que contiene SPADIES, y que analizar estos datos llevará a obtener conclusiones sesgadas con tendencia a altas tasas de deserción y alto tiempo de permanencia para los desertores y graduados. Como alternativa para la investigación, se solicitaron los datos a la Oficina de Registro Académico de la Universidad del Valle, la cual fue suministrada con la estructura que se ilustra en las Figuras A-1, A-2 y A-3.

4.2. Construcción y depuración de base de datos de deserción

Respecto a la definición de Tinto (1989), la interpretación de la deserción queda en manos de los investigadores de acuerdo con el problema de estudio. De esta manera, se define para el trabajo de grado que un estudiante es desertor cuando se encuentra inactivo por dos periodos consecutivos.

Inicialmente se identificó que la base de datos suministrada por Registro académico no indicaba qué estudiantes se encontraban en calidad de desertor, puesto que la base de datos compartida solo tenía información de matrícula y grado. Por lo tanto, no categoriza por estados a los estudiantes, así que a partir de la base de datos compartida se unificó la información entregada de forma general, con el fin de realizar las siguientes categorías para cada estudiante:

- **Activo:** Estudiante que actualmente se encuentra matriculado en el pregrado.
- **Desertor:** Estudiante que se encuentra ausente a partir de 2 o más semestres académicos.

- **Graduado:** Estudiante que finaliza la carrera universitaria.

Seguidamente se agregó la categoría “Num” donde especifica el número de período académico matriculado, el número de semestres correspondiente para cada estudiante ya sea activo, desertor o graduado (mirar Figura A-4).

Una vez desarrollada la categoría “Num”, se intentó construir las tablas de mortalidad considerando todos los semestres de forma individual (no se hizo uso de alguna agrupación entre semestres) realizando el análisis exploratorio de datos. Sin embargo, luego de realizar este ejercicio se encontró periodos en los cuales no existían matrículas de estudiantes, lo anterior, se presenta por la admisión anual y no semestral lo cual no permite tener estudiantes en todos los semestres de evaluación, ejemplo:

Si el presente periodo es de admisión (con la condición que el programa abre matrícula de forma anual), existirán estudiantes matriculados en semestres impares (1,3,5,...) y no estudiantes matriculados en periodos pares, para el siguiente periodo posterior a la admisión existirán estudiantes matriculados en semestres pares (2,4,6,...) y ningún estudiante matriculado en semestres impares.

La dinámica anterior expuesta no será modelable debido a la gran abundancia de falsos valores cero de la deserción, así que, se opta por agrupar los semestres en periodos anuales, y, por ende se construyó una tabla de mortalidad abreviada. En la Tabla 4-1 se presenta la agrupación de los semestres representados por años de la siguiente manera:

Tabla 4-1.: Agrupación de semestres para la construcción de las tablas de mortalidad abreviadas.

Edad	1	2	3	4	5	6	7	8
Rango de semestre	1-2	3-4	5-6	7-8	9-10	11-12	13-14	15>

Luego de tener la base de datos totalmente depurada y organizada, se inició con la construcción de las tablas de mortalidad abreviadas, descartando algunos períodos académicos por lo siguiente:

- Febrero 2019 - mayo 2019 por amnistía académica la cual permite la admisión extraordinaria de estudiantes que hayan completado más del 80 % de su carrera en el pregrado, esto fue aprobado por el consejo superior y reglamentado por el Consejo Académico de la Universidad del Valle.
- Febrero 2020 - mayo 2020 y agosto 2020 - diciembre 2020 debido a que no se alcanzan a generar los dos periodos académicos de inactividad para que un estudiante pueda clasificarse como desertor.

4.3. Cálculo y construcción de tablas de mortalidad abreviadas

Establecidos los periodos de estudio, se definen los siguientes componentes de la tabla de mortalidad, definidos en términos de deserción académica:

- **Edad por períodos académicos (x):** Se estableció que para los años matriculados se agrupaban dos periodos académicos. Esto para los años del 1 al 7 (en la cual se mueve o se describe el 95 % de la población), a partir del año 8 se agruparon los estudiantes que se matricularon de 15 periodos en adelante. Esta decisión se toma debido a que la admisión en el programa de estadística es anual, con lo cual, haciendo seguimiento semestral se encontraran edades (semestres académicos) en los cuales ningún estudiante está matriculado. Por ejemplo, en semestres 2,4,6,8,10, cuando se analiza un periodo académico con admisión de estudiantes, y para semestres 1,3,5,7,9, cuando el periodo académico que se analiza es posterior al periodo de admisión. La situación anterior se resuelve bajo una estrategia de seguimiento anual.
- **Amplitud (n):** Como en el ítem anterior, se establece que para cada año hay dos periodos, por lo tanto, los años del 1 al 7 tienen una amplitud igual a 2 ($n=2$), respecto a la edad de >8 se calcula la mediana de los periodos académicos obteniendo una amplitud igual a 7 ($n=7$).
- P_x : Corresponde al número de estudiantes matriculados para el periodo respectivo.
- d_x : Número de estudiantes que cumplen con la condición de desertor en el año correspondiente (Estudiantes que no registran matrícula durante dos periodos seguidos).
- nm_x : Tasa central de mortalidad académica, la cual se calcula a partir del cociente número de estudiantes en condición de desertor y el número de estudiantes matriculados. (ver Ecuación 3-1).
- nd_x : Número de deserciones de la cohorte ficticia inicial que tiene lugar en períodos entre las edades x y $x + n$.
- nl_x : Representa el número de estudiantes de la cohorte ficticia inicial que llegan activos a la edad x .
- Para el cálculo de nd_x y nl_x se define una cohorte ficticia de 100 estudiantes, la cual asume que en un periodo académico todos los estudiantes ingresaron en forma simultánea. El valor 100 se selecciona al considerar que hasta el año 2014 en el pregrado Estadística el número de estudiantes admitidos eran 100.

Cuando se construye las tablas de mortalidad abreviadas y se valida la información de los periodos académicos para cada edad ($1 a > 8$), se encuentran valores faltantes que generan inconsistencias para realizar las estimaciones de los parámetros del modelo. Una alternativa para completar la información y mitigar este fenómeno fue la imputación de datos, por medio de sustitución de medias, que consiste en asignar un dato o valor ausente por la media de los valores válidos.

Para el proceso de imputación, se tomó un periodo académico faltante teniendo en cuenta la edad, luego se seleccionó los tres periodos anteriores y los tres posteriores correspondientes con la misma edad del periodo faltante, eje: AGOSTO/2012 - DICIEMBRE/2012, Edad 3 (Ver Figura 4-2). El dato imputado corresponde a la media de la tasa de deserción de los 6 periodos, vecinos más cercanos.

GENERAL	PERIODO	Edad	npx	nDx
AGOSTO/2009 - DICIEMBRE/2009	2009-2	3	22	1
AGOSTO/2010 - DICIEMBRE/2010	2010-2	3	24	3
AGOSTO/2011 - DICIEMBRE/2011	2011-2	3	20	2
AGOSTO/2012 - DICIEMBRE/2012	2012-2	3		
AGOSTO/2013 - DICIEMBRE/2013	2013-2	3	27	3
AGOSTO/2014 - DICIEMBRE/2014	2014-2	3	23	3
AGOSTO/2015 - DICIEMBRE/2015	2015-2	3	16	2

Figura 4-2.: Ejemplo de imputación de datos

Realizado el proceso de imputación de datos, se procede a construir la tabla de mortalidad dinámica abreviada presentada en la Figura 4-3. La Tabla para este trabajo de investigación es el objeto de análisis sobre el cual se realizan las labores de Análisis Exploratorio de Datos y ajuste del modelo Lee- Carter que se describe a continuación.

GENERA	PERIODO	Edad	n	npz	nDz	nmz	naz	nqz	lz	dz
ENERO/20	2000-1	1	2	87	24	0,275862	1,5	0,484848	100	10
ENERO/20	2000-1	2	2	29	5	0,172414	1	0,294118	90	20
ENERO/20	2000-1	3	2	23	2	0,086957	1	0,16	70	0
ENERO/20	2000-1	4	2	26	1	0,038462	0	0,071429	70	0
ENERO/20	2000-1	5	2	7		0,025	0	0	70	0
ENERO/20	2000-1	6	2	15		0	0	0	70	0
ENERO/20	2000-1	7	2	7	1	0,142857	1	0,25	70	20
ENERO/20	2000-1	>8	7	12		0,13	0	0	50	0
AGOSTO/20	2000-2	1	2	87	9	0,103448	1,555556	0,197802	100	11
AGOSTO/20	2000-2	2	2	55	12	0,218182	1,166667	0,369231	89	24
AGOSTO/20	2000-2	3	2	11		0,15	0	0	65	4
AGOSTO/20	2000-2	4	2	11		0,12	0	0	61	0
AGOSTO/20	2000-2	5	2	7		0,13	0	0	61	0
AGOSTO/20	2000-2	6	2	13		0,1	0	0	61	0
AGOSTO/20	2000-2	7	2	7	2	0,285714	1,5	0,5	61	0
AGOSTO/20	2000-2	>8	7	14		0,05	0	0	61	0
ENERO/20	2001-1	1	2	53	19	0,358491	2	0,716981	100	9
ENERO/20	2001-1	2	2	49	16	0,326531	1,375	0,542373	91	9
ENERO/20	2001-1	3	2	24	1	0,041667	1	0,08	82	6
ENERO/20	2001-1	4	2	21		0,03	0	0	76	5
ENERO/20	2001-1	5	2	20	1	0,05	0	0,090909	71	0
ENERO/20	2001-1	6	2	7		0,02	0	0	71	0
ENERO/20	2001-1	7	2	16		0,01	0	0	71	0
ENERO/20	2001-1	>8	7	16		0,05	0	0	71	5
AGOSTO/20	2001-2	1	2	62	7	0,112903	1,142857	0,205882	100	10
AGOSTO/20	2001-2	2	2	44	12	0,272727	1,166667	0,444444	90	17
AGOSTO/20	2001-2	3	2	26	2	0,076923	1,5	0,102564	72	4

Figura 4-3.: Tabla de mortalidad dinámica abreviada para la mortalidad académica en el pregrado Estadística de la Universidad del Valle (Periodo)

4.4. Análisis exploratorio de datos

La etapa exploratoria se realizó con el propósito de caracterizar de manera cuantitativa y cualitativa el fenómeno de la deserción al interior del programa académico de Estadística de la Facultad de Ingeniería de la Universidad del Valle. Se realiza un análisis descriptivo del estudio para las variables de Periodo, Edad y el número de estudiantes desertores (d_x), se emplearon las herramientas de gráficos de cajas y de gráficos de representación en líneas de tiempo, con el fin de representar el cálculo de la tasa de deserción global por periodo (representación en líneas de tiempo), el cálculo de la tasa media de deserción por semestre académico (representación en diagramas de cajas y en líneas de tiempo simultáneas para la tasa observada de deserción en cada combinación periodo - semestre). Se realiza un comparativo entre géneros (Masculino y Femenino).

4.5. Ajuste del Modelo Lee-Carter

Con el objetivo de identificar el nivel de deserción en el programa académico Estadística de la Facultad de Ingeniería de la Universidad del Valle, se ajusta el modelo de Lee-Carter.

Para la ejecución del modelo se utiliza el programa R, junto a la librería GNM de la autoría de Turner and Firth (2015), útil para especificar y ajustar modelos no lineales generalizados. Este paquete tiene como uno de sus ejemplos el modelo Lee-Carter el cual se puede ver como un modelo no-lineal generalizado debido a la presencia del término bilineal (Ochoa, 2015).

Inicialmente, se realizó el proceso de estimación del modelo bajo variable respuesta con distribución binomial, según lo planteado por Lee and Carter (1992). Sin embargo, en la estimación de los parámetros este modelo no presentaba convergencia, debido a que en el caso de estudio, se presentaron semestres particulares en los cuales la deserción fue nula (0 % de desertores) o universal (100 % desertores). Estos casos, aunque poco frecuentes, se observan en situaciones relacionadas con cohortes falsas que se conforman cuando un grupo de estudiantes reingresan al programa o ingresan traslado o reingresan, quedando artificialmente ubicados en un periodo académico con muy bajo número de matrículas. Situación similar ocurre en procesos de amnistía.

Como alternativa, y luego de realizar una revisión en la literatura, se encontró que Brouhns et al. (2002) modificó el modelo LC por un modelo log-bilineal de regresión Poisson, concretamente, un modelo no lineal generalizado de tipo Poisson. Por lo tanto, siguiendo el lineamiento del trabajo de maestría de Ochoa (2015), se procedió a usar la distribución Poisson adicionando un término offset del logaritmo de P_x (estudiantes matriculados que se encuentran expuestos a la deserción) en la función de enlace para la ecuación del Modelo de Regresión Poisson.

$$\text{Ln}(d_x) = \text{Ln}(nm_x) + a_x + b_x k_t + \varepsilon_{xt} \quad (4-1)$$

$$d_x \sim \text{Poisson}(\lambda_{xt}) \quad (4-2)$$

Ya que $\text{Ln}(P_x)$ es un componente predicho, la manera de ingresarlo al modelo es a partir del Offset, matemáticamente se observaría de la siguiente manera.

$$\text{Ln}(d_x) = \text{Offset}(\text{Ln}(P_x)) + a_x + b_x k_t + \varepsilon_{xt} \quad (4-3)$$

Donde $\text{Offset}(\text{Ln}(P_x))$ es la reconfiguración, $a_x + b_x k_t$ es la parte estimable del modelo y por último ε_{xt} es el error aleatorio.

En un principio, se definen estimaciones iniciales con la intención de contribuir a la convergencia del proceso de estimación. Estableciéndose punto inicial con $b[1] = 1$ y $k[1] = 0$, donde se tiene 8 edades y 40 periodos (años). Por tanto, se podría afirmar que al correr el programa con valores iniciales aleatorios y llegar después a transformaciones a un mismo vector de parámetros significa que, con valores de tolerancia y número máximo de iteraciones, sí se llega a un mínimo global (Currie, 2014).

Siguiendo la sintaxis para estimar un modelo, según Ochoa (2015) consiste en definir las variables Periodo, Edad, nm_x , P_x , contenidas en un data.frame, por ejemplo, Data, con la siguiente instrucción en el programa R:

```
m1 = gnm(nm_x Edad + Mult(Edad, Periodo), offset = log(P_x),
```

```
family = "poisson", data = Data)
```

- **fórmula :** (Ej. nm_x Edad + Mult(Edad,Periodo)), Fórmula en la cual se indica la relación entre los parámetros a estimar y el dato observado. En este caso será nm_x (Tasa central de Deserción) se regresará contra un término multiplicativo en Periodo y Edad.
- **offset:** Este término indica que hay información adicional que debe incluirse en el análisis. En este caso serán los expuestos al riesgo (P_x) estudiantes matriculados.
- **family:** Especifica la función de enlace del modelo. Esta opción permite utilizar otras familias de modelos exponenciales, como “quasi-poisson” con la cual se estima un modelo Poisson con parámetro de sobre-dispersión ϕ Ochoa (2015).

Para obtener una única solución del modelo se realiza la normalización de los parámetros.

Luego de realizar el análisis exploratorio de datos y el ajuste del modelo Lee-Carter, se realizan comparaciones de los resultados del modelo discriminado por hombre y mujer, donde se logra establecer los patrones temporales y el comportamiento estructural para cada uno. Finalmente, se verifica el ajuste del modelo de forma general y por género, por medio del análisis de residuos representados en diagramas de cajas para las 8 edades y 40 periodos que componen el estudio.

5. Resultados

En esta sección se presentan los resultados de la investigación. En la primera parte se muestra el análisis descriptivo de las variables de estudio, utilizando diagrama de cajas y gráficos de tendencia. Inicialmente se describe el comportamiento de forma general para la población objeto de estudio, y, seguidamente se discrimina el comportamiento de la deserción por género. Para el desarrollo del modelo Lee carter, se utilizó un modelo de regresión lineal no generalizado a partir de una función LogPoisson perteneciente a la familia exponencial la cual permitió la estimación de los tres parámetros. Finalmente, se compara el comportamiento estructural y la tendencia de la deserción entre hombres y mujeres.

5.1. Análisis Exploratorio

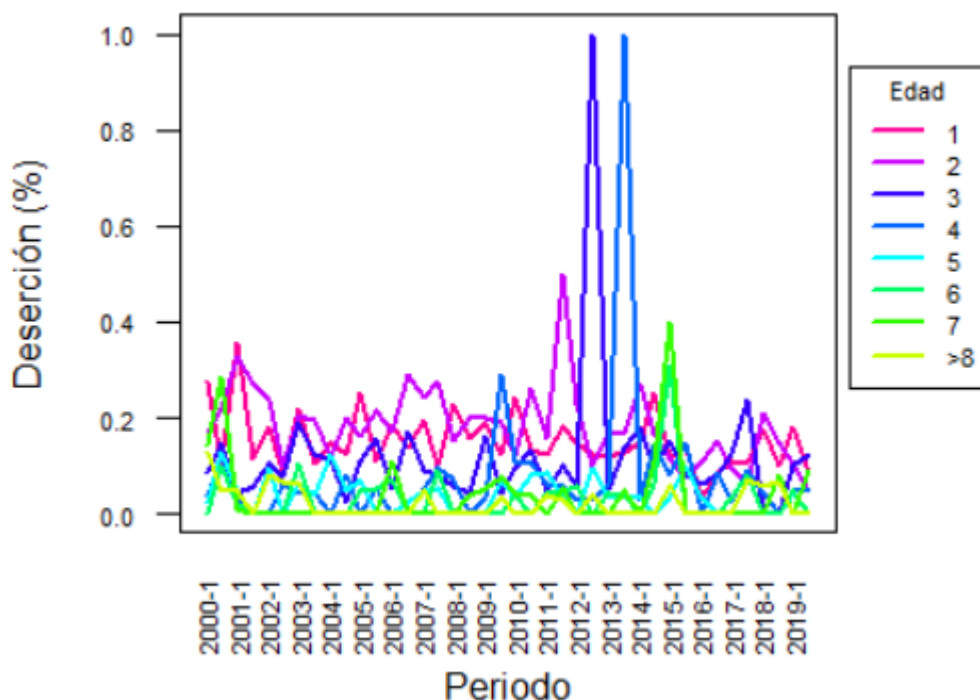


Figura 5-1.: Porcentaje de deserción de estudiantes del programa académico Estadística de la Universidad del Valle

En la Figura 5-1 se muestra la deserción general (simplificado gráficamente en 20 años), a partir del año 2000-1 hasta el periodo 2019-2, para los 40 periodos estudiados, discriminado por líneas de colores para las 8 edades. De acuerdo con esta información, se presentan dos picos pronunciados con un porcentaje de deserción de un 100 %, correspondientes a las edades 3 (matrícula de semestre 5 y 6) y 4 (matrícula de semestre 7 y 8), para los periodos 2012-2 y 2013-2, respectivamente. De esta forma, se evidencia que para los periodos nombrados anteriormente hubo algún tipo de novedad. Probablemente relacionadas con la presencia de un grupo conformado por estudiantes de traslado o reingreso, que se ubicaron en estos periodos académicos y posteriormente desertaron.

Además, en las edades iniciales (1 y 2) el porcentaje de deserción a lo largo de los periodos se observa un aumento importante en la deserción estudiantil, frente a las edades medias (3, 4 y 5) y las edades mayores (7 y >8), esta última se concentra en un rango porcentual de 0 % a 10 %. Quiere decir que, se presenta una disminución importante en el porcentaje de deserción general cuando los estudiantes están prontos a culminar sus estudios.

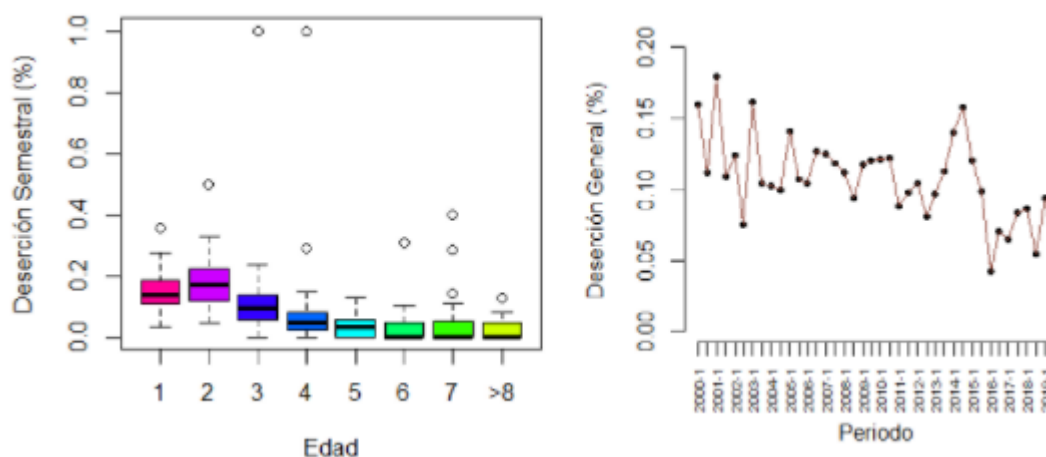


Figura 5-2.: Porcentaje de deserción de estudiantes pregrado estadística Universidad del Valle por periodo y edad

Para la Figura 5-2 se realizó un diagrama de cajas, donde se puede visualizar el porcentaje de deserción respecto a la edad diferenciando por colores cada edad correspondiente. En esta gráfica se puede observar los puntos atípicos extremos en las edades 3 y 4, ya que se encuentran bastante alejados de la media. Respecto a la edad 7 es aquella que tiene más datos fuera del límite superior, es decir tiene datos extremos leves. En general, la deserción respecto a la edad no tiene una gran dispersión, sin embargo, a partir de la edad 4 las cajas son más angostas lo que indica que a medida que los estudiantes ingresan a semestres superiores (aumento de edad) la puntuación de dispersión va disminuyendo, y a su vez el porcentaje de deserción se acerca a valores cercanos al 0 %.

Por otro lado, respecto a la tendencia temporal se resalta que a lo largo de los 20 años que acoge el estudio (40 periodos) se visualiza una leve inclinación decreciente. Inicialmente, el porcentaje de deserción para el periodo 2000-1 es del 16,01 %, y a medida que pasa el tiempo para el año 2019-2 se observa un porcentaje de deserción general del 5,92 %. A excepción de los periodos del 2000-1 (17,61 %), del 2003-1 (16,13 %) y del 2014-2 (15,75 %).

Como el objetivo principal de este proyecto, es analizar a comportamiento estructural y la tendencia de la deserción en el programa académico de Estadística, se hace indispensable realizar el contraste exploratorio de la deserción discriminado por género. De esta forma, se logra tener un panorama descriptivo la mayor deserción a lo largo del tiempo en hombres y mujeres.

5.1.1. Análisis Exploratorio Mujeres

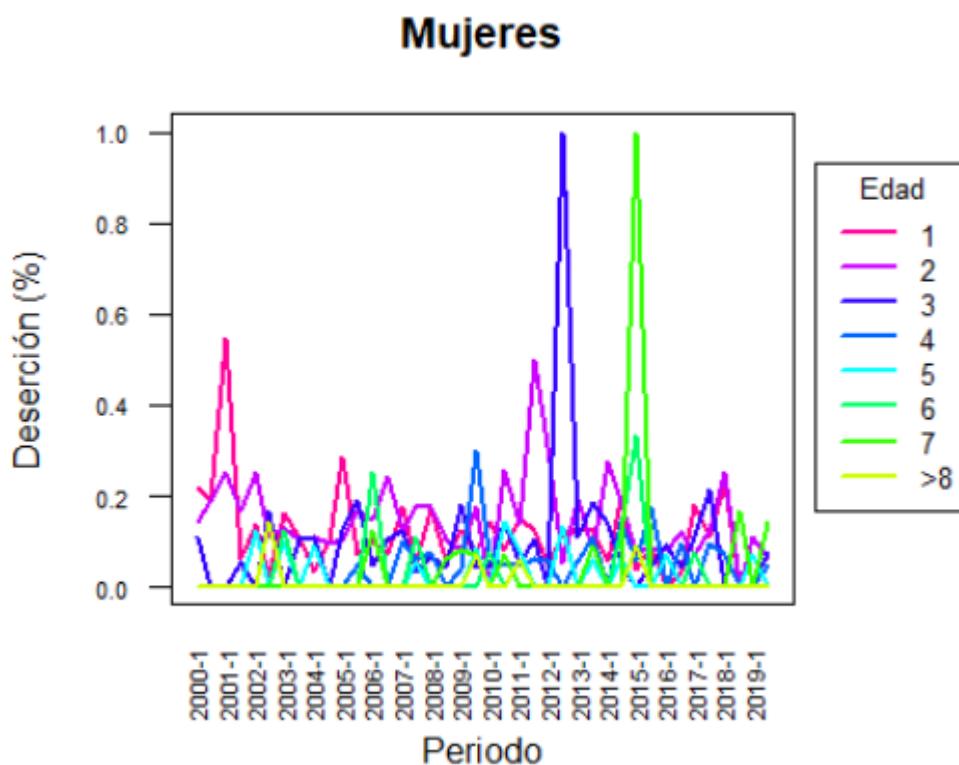


Figura 5-3.: Porcentaje de deserción de estudiantes mujeres del programa académico Estadística de la Universidad del Valle

Realizando una comparación entre la figura 5-1 y la figura 5-3 se observa que el crecimiento en la deserción general en las edades iniciales se le atribuye principalmente a los altos porcentajes de deserción de las mujeres en las edades 1 y 2. Adicionalmente, se observan

los picos con un porcentaje de deserción del 100 % en las edades 3 (semestre matriculado 5 y 6) para el periodo 2012-2 y la edad 7 (semestre matriculado 13 y 15) en los periodos 2015-1.

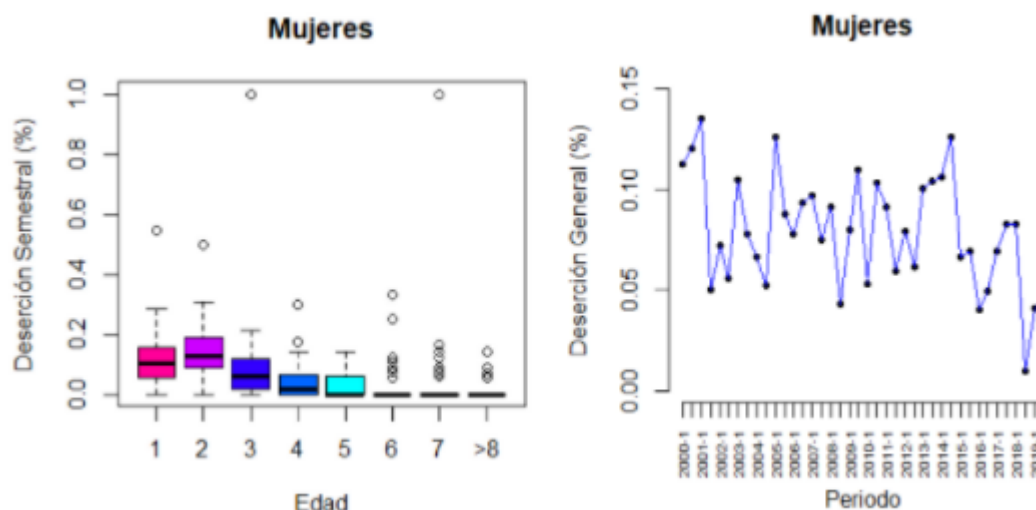


Figura 5-4.: Porcentaje de deserción de las estudiantes mujeres del programa académico Estadística de la Universidad del Valle por periodo y edad

En la Figura 5-4 observamos la gráfica de cajas donde se visualiza dos puntos atípicos extremos en las edades 3 y 7 adicional el porcentaje de deserción a partir de la edad 6 tiende a ser nulo, a excepción de algunos datos atípicos alejados de la media. Al igual que la deserción general, representada en la figura 6.2 para la edad 2 (semestre matriculado 3 y 4) se aprecia que el valor medio es superior al valor de las edades restantes puesto que el porcentaje medio de deserción equivale a un 18 %.

Ahora, con el porcentaje de deserción frente a los periodos 2000-1 y 2014-2 el porcentaje de deserción es inestable y no tiene tendencia, hasta que llega al periodo 2015-2, donde se evidencia que a medida que el tiempo avanza la deserción en mujeres tiende a decrecer al tal punto de tomar valores cercanos 0 %.

5.1.2. Análisis Exploratorio Hombres

Realizando una comparación entre las Figuras 5-1 y 5-5 se puede observar que la deserción en los hombres para las edades iniciales 1 (semestre matriculado 1 y 2) y 2 (semestre matriculado 3 y 4) parece variar relativamente de manera constante, no toma valores extremos todo lo contrario a la deserción en mujeres en estas mismas edades. Adicionalmente tiene una deserción del 100 % en las edades correspondientes a las edades 3 (matrícula de semestre 5 y 6) y 4 (matrícula de semestre 7 y 8), para los periodos 2012-2 y 2013-2.

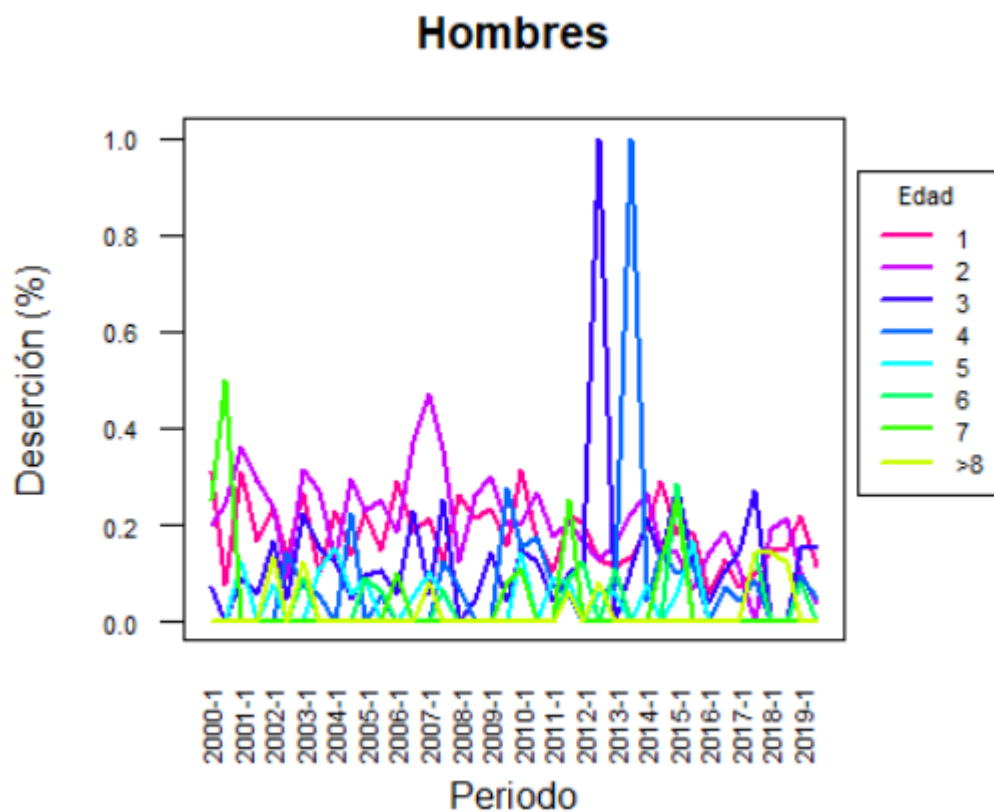


Figura 5-5.: Porcentaje de deserción de estudiantes hombres del programa académico Estadística de la Universidad del Valle

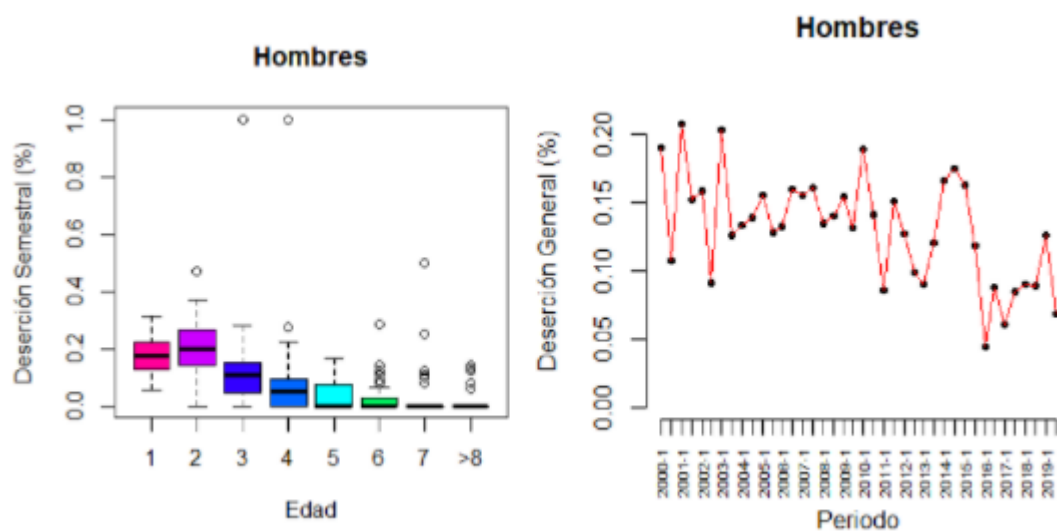


Figura 5-6.: Porcentaje de deserción de las estudiantes hombres del programa académico Estadística de la Universidad del Valle por periodo y edad

En la Figura 5-6, respecto al porcentaje de deserción por edad se observa dos puntos atípicos extremos en las edades 3 y 4. Adicionalmente, se visualiza que en la edad 2 se concentra un alto porcentaje de deserción entre el 18 % al 25 % aproximadamente, con una media del 20 %. Sin embargo, a medida que la edad aumenta el porcentaje de deserción se acerca al 0 %, a excepción de los puntos atípicos leves que se encuentran a partir de la edad 6.

Para observar el comportamiento temporal, aunque se evidencia que hay tendencia decreciente, comparado con las Figuras 5-2 y 5-4, se visualiza que esta tendencia no es muy marcada para los hombres. No obstante, para los periodos del 2000-1 hasta 2004-2, es observable que en las cohortes iniciales de cada año representadas con “1” suele tener mayor porcentaje de deserción que en las cohortes a mitad de año “2” para cada año respectivo. Caso contrario ocurre para los periodos centrales (2005-1 hasta 2011-2), no se destaca un máximo de deserción, considerando que para estos periodos no se observa mayor variabilidad.

5.2. Ajuste del Modelo Lee Carter

En la Figura 5-7 se ilustran las estimaciones de los parámetros para el modelo de Lee-Carter representado en la Ecuación 3-4 para población en general de los estudiantes de Estadística de la Universidad del Valle, conforme al uso de la librería GNM en el programa R. En el análisis del parámetro a_x , la tendencia promedio de la deserción se visualiza en el panel I, se detecta un comportamiento descendente en el gráfico (línea sólida-spline). Se trata de un máximo global para las edades 1 y 2 que corresponde, a un incremento del número de deserciones y se asocia con los semestres de primero a cuatro en el cual el estudiante inicia su preparación universitaria. Por otro lado, se observa que a medida que aumenta la edad, la deserción desciende de manera rápida para las edades 6, 7 y 8. Esto último, confirma un comportamiento habitual (mirar Figura 5-2) para los semestres superiores puesto que se presenta menor deserción.

En el panel II se observa la estimación del parámetro k_t , donde se muestra la tendencia temporal que ha tenido la deserción a nivel general desde el año 2000-1 hasta el 2019-2. En un inicio, se presenta un incremento de la deserción desde el periodo 2000-1 hasta 2004-2. Seguidamente, empieza a disminuir a tal punto de considerarse estable en los periodos 2005-1 hasta el 2014-2. Finalmente, para el periodo 2015-1 se observa nuevamente un incremento en el índice de deserción de estudiantes.

Por último en el panel III, se visualiza la interacción edad-periodo ($b_x k_t$) que indica el comportamiento de la mortalidad académica por cada edad. Como se puede apreciar en la Figura 5-7, se permite inferir que durante la ventana de observación de la deserción universitaria en el pregrado Estadística de la Universidad del Valle; para las edades 1 y 2, b_x toma valores negativos, lo que indica que en estas edades la deserción incrementa al variar el tiempo, respecto a las edades 3, 4 y 5 el parámetro, toma valores cercanos a cero, quiere decir

que en estas edades la deserción se puede considerar estable. Finalmente, para las edades 6,7 y 8 la estimación toma valores positivos lo que significa una disminución de la mortalidad académica a través del tiempo.

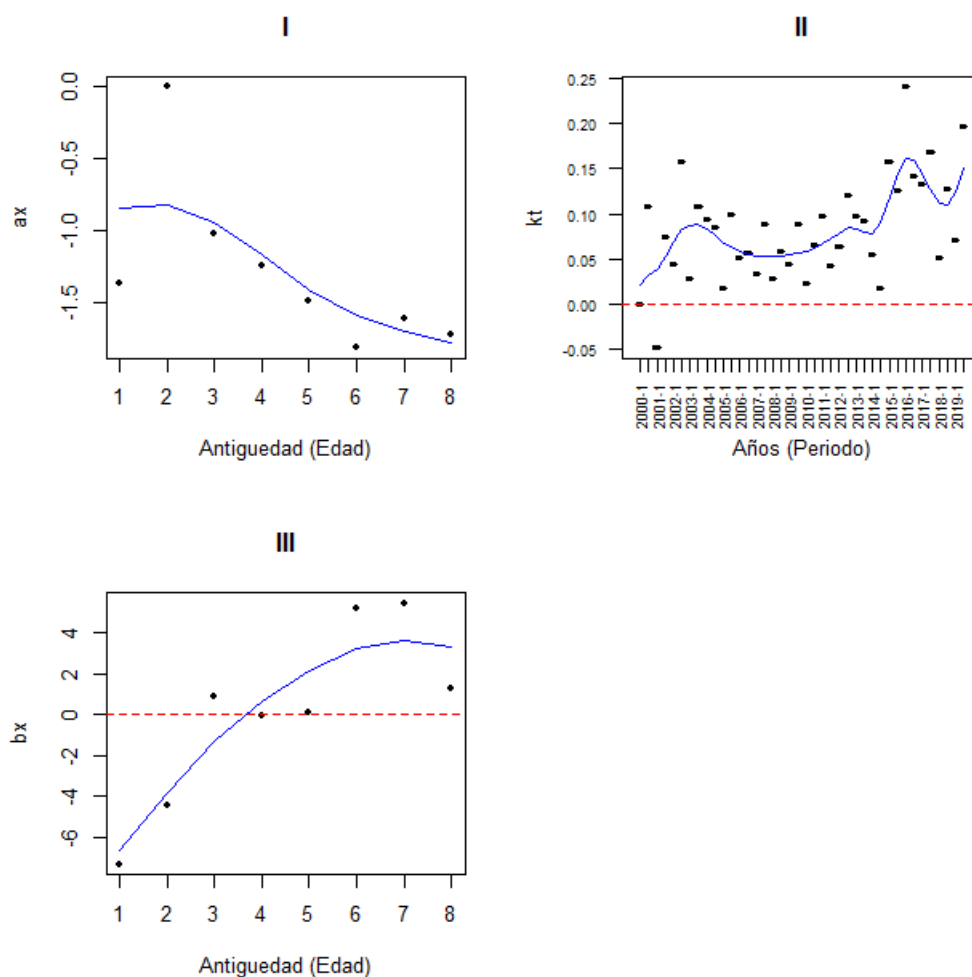


Figura 5-7.: Estimaciones de parámetros a_x , b_x y k_t del modelo Lee Carter. Datos generales

5.2.1. Ajuste del Modelo Lee Carter- Mujeres

En la Figura 5-8, se observan las estimaciones de los parámetros a_x , k_t y b_x para el modelo Lee-Carter desarrollado para la base de datos de mujeres. En el panel I se visualiza el comportamiento del parámetro a_x , inicialmente se forma una línea ascendente hasta la edad 3 donde se encuentran los semestres 5 y 6, teniendo un máximo global para la edad 2, que comprende a los semestres 3 y 4. Así que, para las edades nombradas la mortalidad académica en las mujeres matriculadas en el programa académico de estadística es más alta hasta llegar a la edad 4. Donde, el comportamiento del fenómeno de la deserción empieza

a descender a medida que aumenta la edad, salvo en la edad 7 donde se observa que las mujeres matriculadas en los semestres 13 y 14, suelen tener las mismas características de la edad 3.

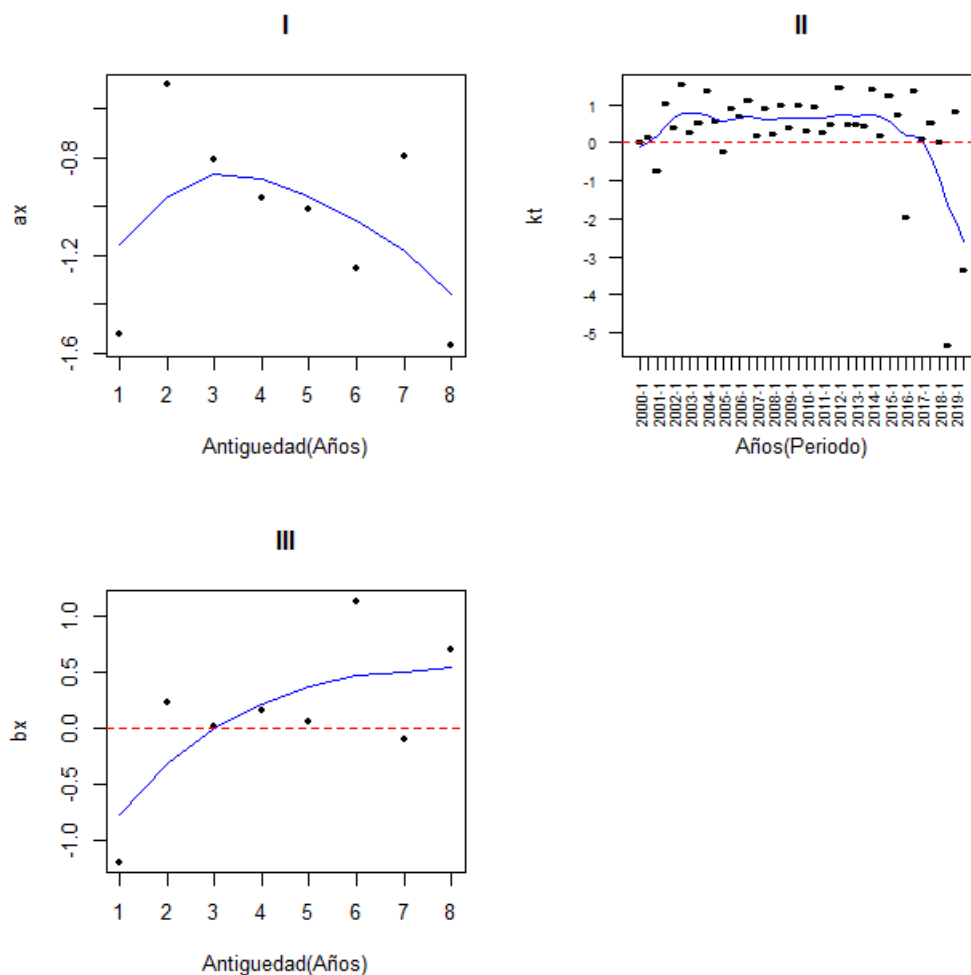


Figura 5-8.: Estimaciones de parámetros a_x , b_x y k_t del modelo Lee Carter. Mujeres

En el panel II se muestra la tendencia temporal de k_t donde la deserción ha permanecido estable salvo los últimos tres periodos que se encuentran por debajo de la línea de corte cero, asociados a los años 2016-1, 2018-2 y 2019-2 con sus cohortes correspondientes, en el que se observa una reducción de la deserción. Finalmente, para el panel III con la estimación del parámetro b_x donde la gran mayoría de los puntos están cerca a cero. Sin embargo, el parámetro tiene una forma ascendente donde se observa en el año 1 un aumento en la deserción de mujeres y a medida que el tiempo avanza la deserción va disminuyendo al tomar valores positivos.

5.2.2. Ajuste del Modelo Lee Carter- Hombres

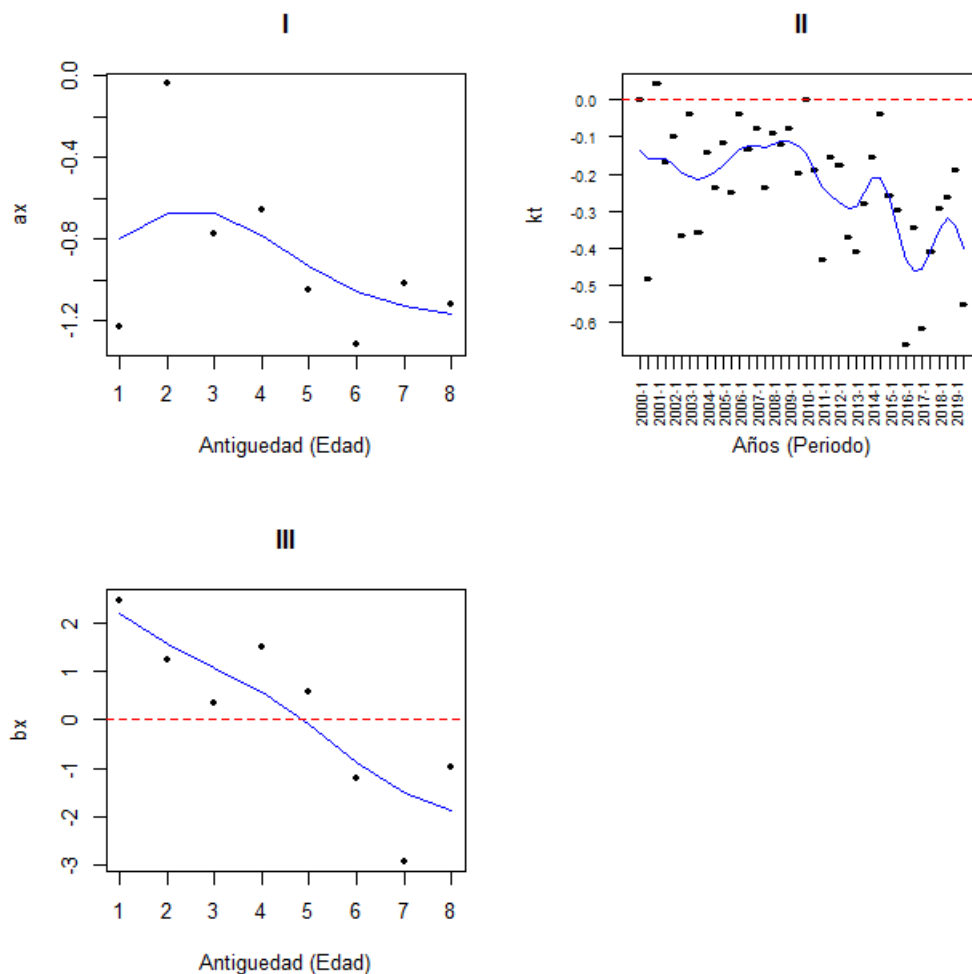


Figura 5-9.: Estimaciones de parámetros a_x , b_x y k_t del modelo Lee Carter. Hombres

En el caso de los hombres, en la Figura 5-9 se representa en el panel I la estimación del parámetro a_x que se evidencia el abandono universitario de los estudiantes en el programa académico de estadística es mayor para la edad 2 correspondiente a los semestres 3 y 4. Por otro lado, para las edades 6, 7 y 8 de forma gráfica tiende a estabilizarse al igual que el análisis general en la Figura 5-7. En el panel II se muestra el parámetro k_t , generalmente tiene una forma decreciente puesto que tiende a disminuir a través del tiempo. En particular, a partir del periodo 2009-1 hasta el año 2019-2 la deserción decrece de forma pronunciada en contraste con los años inferiores al 2009.

Por último, la estimación del parámetro b_x se encuentra en el panel III donde conserva un comportamiento inverso al analizado en la Figura 5-8, quiere decir que los hombres en la

edad 1 hasta la 5 disminuye la deserción a través del tiempo, y, a partir de la edad 6 hasta la 8 el fenómeno de la deserción aumenta, principalmente para la edad 7 correspondiente a los semestres matriculados 13 y 14. En otras palabras, se permite intuir que los hombres a mayor número de semestres matriculados no culminan sus estudios universitarios.

5.2.3. Ajuste del Modelo Lee Carter- Contraste Hombres vs Mujeres

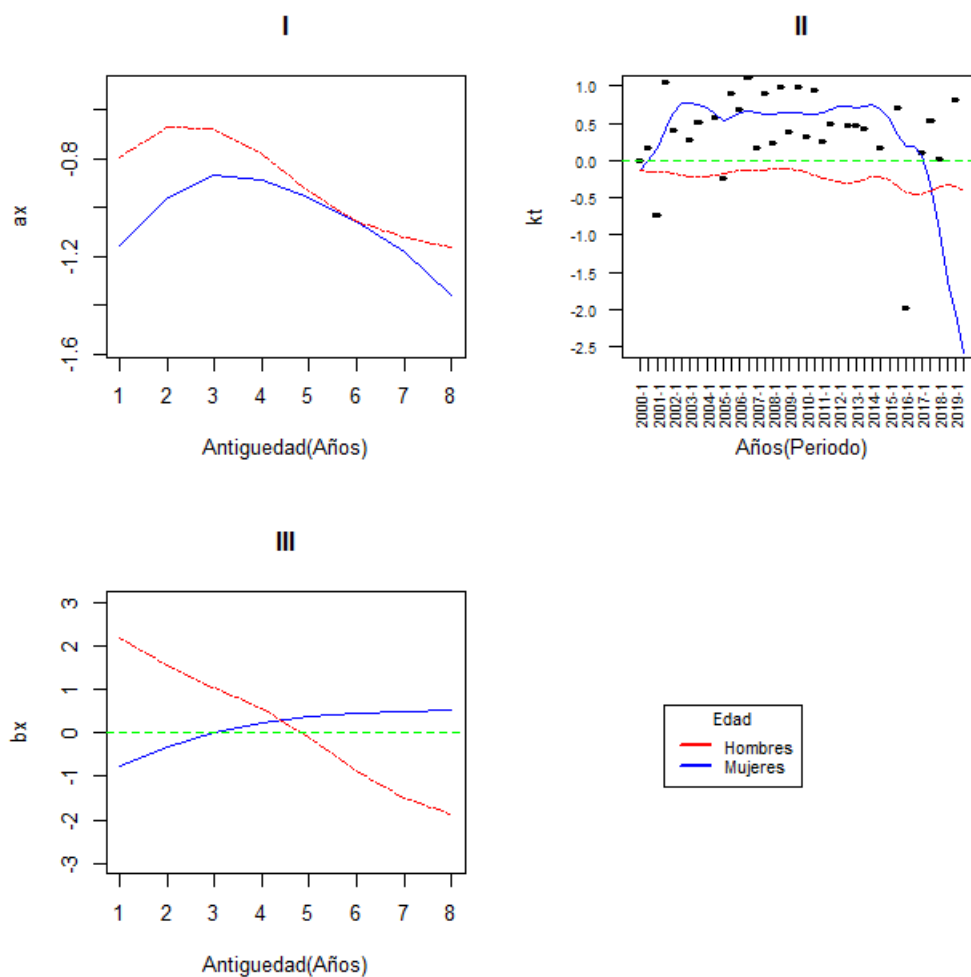


Figura 5-10.: Estimaciones de parámetros a_x , b_x y k_t del modelo Lee Carter. Contraste Hombres Vs Mujeres

Mostrando en un mismo plano las diferencias entre hombres y mujeres, se generó la Figura 5-10, en la cual se observa que la tendencia general de la deserción (a_x) es muy similar. Pero, en la dinámica temporal la reducción de la deserción es diferente entre hombres y mujeres.

Por lo tanto, el comportamiento de la deserción académica con el componente de la edad (b_x) al incluir el componente del tiempo (k_t), su forma es inversa, se exalta en la línea roja (representación de los hombres) que cuando se incorpora el factor del tiempo y la edad, aumenta la deserción en edades superiores a 5. En cuanto a la línea azul (representación de las mujeres) ocurre lo contrario, hay aumento de la deserción para las edades 1 y 2, y va disminuyendo a partir de estas.

5.3. Análisis de Residuos

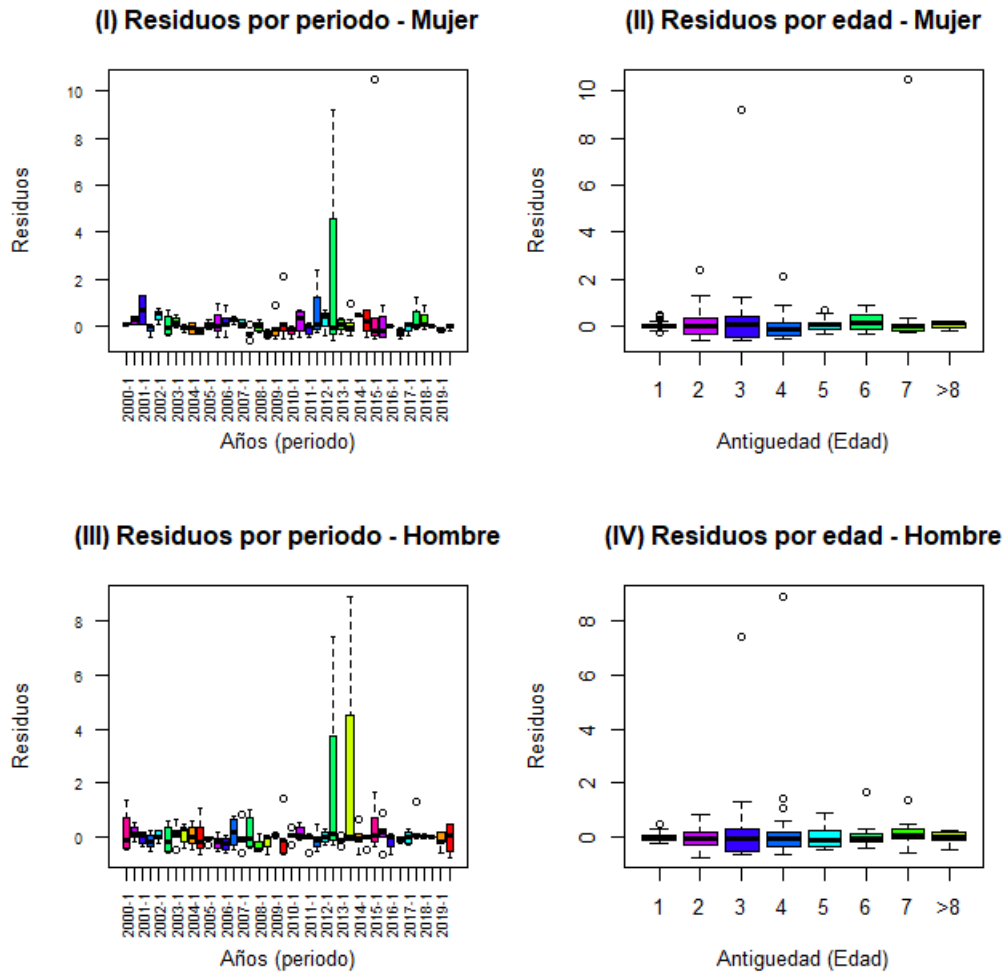


Figura 5-11.: Residuos del modelo Lee-Carter por género, periodo y edad

En la Figura 5-11 se encuentran los residuos por medio boxplot que se representan por género, periodo y edad del modelo Lee-Carter para la dinámica de la deserción universitaria. Los periodos donde se encuentran más variabilidad de los residuos para las mujeres es el año

2012-2, respecto a los hombres la variabilidad se aloja en los años 2012-2 y 2013-2. Por otra parte, los residuos para la edad en ambos géneros presentan mayor dispersión para la edad 3, sin embargo, esta situación no es tan pronunciada respecto a las edades restantes.

6. Conclusiones y recomendaciones

6.1. Conclusiones

Una de las limitaciones para el desarrollo de este proyecto se desencadena principalmente en los datos. Inicialmente, se solicitó y suministró la información por medio del Ministerio de Educación de Colombia, correspondientes al SPADIES. En el proceso de validación de la base de datos, se encontraron inconsistencias en el género de los estudiantes, no hay forma de identificar el instante exacto de tiempo que presenta la condición de desertor, la información reportada por el sistema de información no se encuentra actualizada con lo reportado por Registro académico de la Universidad del Valle.

Con los datos suministrados por Registro Académico de la Universidad del Valle, se logró elaborar las tablas dinámicas de mortalidad, para el análisis de la deserción académica, resolviendo las fallas estructurales de la base de datos del SPADIES. Esta información suministrada por la universidad, para los procesos de análisis de datos internos se considera óptimos en las condiciones actuales, sin embargo, para el desarrollo de la investigación fue necesario realizar algunos ajustes a la base de datos, principalmente la categorización del estado actual del estudiante; atribuidos a estudiante desertor, activo y graduado.

Se identificó en el proceso de ajuste y estimación del modelo Lee-Carter en la población en general, un comportamiento descendiente de la deserción a través del tiempo, es decir, en los semestres superiores la mortalidad académica disminuye.

Estos resultados se ha presentado en la dinámica temporal de la deserción que en los primeros semestres (1° a 4°), los hombres tienen una reducción de la mortalidad académica a lo largo del tiempo. Caso contrario ocurre con las mujeres, donde muestra un incremento en la deserción académica. Sin embargo, la deserción académica de las mujeres es menos cambiante que la presentada en los hombres, en específico para las edades de 3 a 5.

Como investigadoras de este proyecto suponemos que el comportamiento de la deserción en los hombres ocurre debido a hay tendencia a que los estudiantes ingresen a la universidad en edad promedio de 17 a 20 años, y a medida que van desarrollando sus estudios tienen a tener responsabilidades económicas que los fuerzan a abandonar sus estudios, esto con base al estudio realizado por Simonds (2012) que concluye que el riesgo de desertar en los hombres es de 1.7 mayor que el de las mujeres.

6.2. Recomendaciones

Se sugiere que se pueda extender el estudio a toda la facultad de Ingeniería, puesto que al analizar sólo un programa académico se tiene la limitante de no contar con información completa. También, se hace importante conocer y comparar la deserción en los diferentes programas académicos.

Adicionalmente, para futuros estudios, se puede realizar la proyección de la deserción en el programa académico Estadística de la Universidad del Valle a partir del modelo Lee-carter, lo cual puede ayudar a disminuir la alta mortalidad académica para los primeros semestres.

Bibliografía

- Amaya Jiménez, L. (2018). Criterio de akaike para la selección de modelos con transformaciones.
- Arriaga, E. E. (1996). Los años de vida perdidos: su utilización para medir los niveles y cambios de la mortalidad. *Notas de población*.
- Bean, J. P. (1985). Interaction effects based on class level in an explanatory model of college student dropout syndrome. *American educational research journal*, 22(1):35–64.
- Brouhns, N., Denuit, M., and Vermunt, J. K. (2002). A poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and economics*, 31(3):373–393.
- Castaño, E., Gallón, S., Gómez, K., Vásquez, J., et al. (2008). Análisis de los factores asociados a la deserción estudiantil en la educación superior: un estudio de caso. *Revista de educación*.
- Cayuela, L. (2009). Modelos lineales generalizados (glm). *Materiales de un curso del R del IREC*.
- Consejo, S. (1997). Acuerdo 009, capítulo viii: Bajos rendimientos académicos. *Cali, Universidad del Valle, Ciudad Universitaria Meléndez, Colombia*.
- Currie, I. D. (2014). On fitting generalized linear and non-linear models of mortality. *Scandinavian Actuarial Journal*, 2016(4):356–383.
- Debón, A., Montes, F., and Sala, R. (2009). Tablas de mortalidad dinámicas. una aplicación a la hipoteca inversa en espana. fundación ico. *Publicaciones de la Universitat de Valencia, valencia*.
- Diaz Rojo, G. and Debón Aucejo, A. M. (2016). Tendencias y comportamiento de la mortalidad en colombia entre 1973 y 2005. *Estadística española*, 58(191):277–300.
- Educación, M. (2016). Reporte sobre deserción y graduación en educación superior año 2016. *Mini educación*. Recuperado de: https://www.mineduacion.gov.co/sistemasdeinformacion/1735/articles-357549_recurso_5.pdf.

- Escobar, J., Largo, E., and Pérez, C. A. (2006). Factores asociados a la deserción y permanencia estudiantil en la universidad del valle (1994-2006). *Cali, Universidad del Valle, Vicerrectoría Académica, Facultad de Ciencias Sociales y Económicas, Centro de Investigación y documentación socioeconómica-CIDSE*.
- Himmel, E. (2002). Modelo de análisis de la deserción estudiantil en la educación superior. *Calidad en la Educación*, (17):91–108.
- Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting us mortality. *Journal of the American statistical association*, 87(419):659–671.
- McCullagh, P. and Nelder, J. (1989). Generalized linear models [ivq].
- Morales, A. F. (2009). Graduación de la mortalidad en andalucía con modelos de mortalidad con heterogeneidad inobservable. In *Anales del Instituto de Actuarios españoles*, number 15, pages 23–50. Instituto de Actuarios Españoles.
- Mosquera Daza, L. A. and Valdés Velásquez, L. J. (2008). Estudio de permanencia de los estudiantes de los programas tecnológicos de la facultad de ingeniería de la universidad del valle, sede meléndez, para el periodo 2002-2006.
- Ochoa, C. (2015). El modelo lee-carter para estimar y pronosticar mortalidad: una aplicación para colombia. *Universidad Nacional de Colombia, sede Medellín*.
- Oficina de Planeación, y. D. I. (2011). Anuario estadístico universidad del valle 2010. *Cali, Universidad del Valle, Rectoría, Oficina de Planeación y Desarrollo Institucional, Ciudad Universitaria Meléndez*.
- Ortiz, F., Villegas, M., and Zarruk, A. (2012). Tablas de mortalidad (mortality tables). *Documentos de Matemáticas y Estadística*, (5).
- Prentice, R. L. and Gloeckler, L. A. (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*, pages 57–67.
- Ruiz, C. G., Muriel, D. M. D., Gallego, J. F., Velez, E. C., Gomez, S. G., and Portilla, K. G. (2009). *Deserción Estudiantil en la Educación Superior Colombiana: Metodología de seguimiento, diagnóstico y elementos para su prevención*. Ministerio de educación nacional.
- Simonds, Jennifer. Silva, P. (2012). Análisis de la deserción de los estudiantes del programa de estadística de la universidad del valle para el periodo 1999-2011.
- Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of educational research*, 45(1):89–125.
- Tinto, V. (1989). Definir la deserción: una cuestión de perspectiva. *Revista de educación superior*, 71(18):1–9.

- Turner, H. and Firth, D. (2015). Generalized nonlinear models in r: An overview of the gnm package. r package version 1.0-8.
- Uribe Gil, J. M. (2017). Tablas de vida de santiago de cali: Tendencias recientes y proyecciones: 1985-2030. *Sociedad y Economía*, (32):171–194.

A. Anexo: Estructura de base de datos de la Oficina de Registro Académico de la Universidad del Valle

B	C
codigo periodo academico	Nombre periodo academico
200208041	AGOSTO/2002 - DICIEMBRE/2002
200301051	ENERO/2003 - JUNIO/2003
200001041	ENERO/2000 - MAYO/2000
200008041	AGOSTO/2000 - DICIEMBRE/2000
200101041	ENERO/2001 - MAYO/2001
200108051	AGOSTO/2001 - ENERO/2002
200203041	MARZO/2002 - JULIO/2002
200301051	ENERO/2003 - JUNIO/2003
200308041	AGOSTO/2003 - DICIEMBRE/2003
200401041	ENERO/2004 - MAYO/2004
200908041	AGOSTO/2009 - DICIEMBRE/2009
201002041	FEBRERO/2010 - JUNIO/2010
201008041	AGOSTO/2010 - DICIEMBRE/2010
200008041	AGOSTO/2000 - DICIEMBRE/2000
200001041	ENERO/2000 - MAYO/2000
200008041	AGOSTO/2000 - DICIEMBRE/2000
200101041	ENERO/2001 - MAYO/2001
200108051	AGOSTO/2001 - ENERO/2002
200203041	MARZO/2002 - JULIO/2002
200208041	AGOSTO/2002 - DICIEMBRE/2002
200301051	ENERO/2003 - JUNIO/2003
200408041	AGOSTO/2004 - DICIEMBRE/2004
201502041	FEBRERO/2015 - JUNIO/2015
201508041	AGOSTO/2015 - DICIEMBRE/2015
200208041	AGOSTO/2002 - DICIEMBRE/2002

Figura A-1.: Estructura de datos de Registro Académico de la Universidad del Valle de código de estudiantes y periodos matriculados

Estudiante	Fecha de Grado
215765	17/11/2000
218135	17/11/2000
215254	23/02/2001
211433	11/05/2001
010589	24/08/2001
010646	24/08/2001
160003	23/11/2001
427945	26/04/2002
430288	14/06/2002
422984	29/11/2002
431217	29/11/2002
219710	29/11/2002
430853	29/11/2002
429696	29/11/2002
517971	29/11/2002
013768	23/05/2003
810994	23/05/2003
911280	23/05/2003
425513	23/05/2003
812543	23/05/2003
015826	5/09/2003
518323	14/11/2003

Figura A-2.: Estructura de datos de Registro Académico de la Universidad del Valle de código de estudiantes y fecha de grado

B	C	D
sexo	Fecha Nacimiento	Estrato
F	1/01/1955	0
M	1/01/1970	0
M	17/11/1957	0
F	1/01/1967	03
M	1/01/1963	0
M	29/06/1969	0
M	1/06/1972	0
M	2/01/1958	0
M	1/01/1968	0
M	3/04/1973	0
M	1/01/1966	0
M	1/01/1972	0
M	25/05/1963	0
M	4/08/1970	0
F	11/10/1971	0
F	29/03/1973	0
F	23/05/1965	0
M	1/01/1955	0

Figura A-3.: Estructura de datos de Registro Académico de la Universidad del Valle de código de estudiantes, sexo, fecha de nacimiento y estrato

Año	codigo periodo	Nombre periodo academico	Num	EstadoDin
2002	200208041	AGOSTO/2002 - DICIEMBRE/2002	1	Activo
2003	200301051	ENERO/2003 - JUNIO/2003	2	Deserción
2000	200001041	ENERO/2000 - MAYO/2000	1	Activo
2000	200008041	AGOSTO/2000 - DICIEMBRE/2000	2	Activo
2001	200101041	ENERO/2001 - MAYO/2001	3	Activo
2001	200108051	AGOSTO/2001 - ENERO/2002	4	Activo
2002	200203041	MARZO/2002 - JULIO/2002	5	Activo
2003	200301051	ENERO/2003 - JUNIO/2003	6	Activo
2003	200308041	AGOSTO/2003 - DICIEMBRE/2003	7	Activo
2004	200401041	ENERO/2004 - MAYO/2004	8	Activo
2009	200908041	AGOSTO/2009 - DICIEMBRE/2009	9	Activo
2010	201002041	FEBRERO/2010 - JUNIO/2010	10	Activo
2010	201008041	AGOSTO/2010 - DICIEMBRE/2010	11	Graduado
2000	200008041	AGOSTO/2000 - DICIEMBRE/2000	1	Deserción
2000	200001041	ENERO/2000 - MAYO/2000	1	Activo
2000	200008041	AGOSTO/2000 - DICIEMBRE/2000	2	Activo
2001	200101041	ENERO/2001 - MAYO/2001	3	Activo
2001	200108051	AGOSTO/2001 - ENERO/2002	4	Activo
2002	200203041	MARZO/2002 - JULIO/2002	5	Activo
2002	200208041	AGOSTO/2002 - DICIEMBRE/2002	6	Activo
2003	200301051	ENERO/2003 - JUNIO/2003	7	Activo
2004	200408041	AGOSTO/2004 - DICIEMBRE/2004	8	Activo

Figura A-4.: Estructura de datos con categorías para cada estudiante