



Universidad Veracruzana

Facultad de Estadística e Informática

Región Xalapa

Especialización en Métodos Estadísticos

Métodos de estimación en regresión de Cox para calcular la
probabilidad de supervivencia académica en estudiantes
universitarios

Reporte de aplicación para obtener el diploma de Especialista en
Métodos Estadísticos

Presenta:

Isabel Vega Rebolledo

Director:

Dr. José Juan Muñoz León

Codirectora:

Dra. Cecilia Cruz López

Asesor:

Dr. Ángel Juan Sánchez García

Octubre de 2025

“Lis de Veracruz: Arte, Ciencia, Luz”





Universidad Veracruzana



Universidad Veracruzana

Facultad de Estadística e Informática
Región Xalapa

Especialización en Métodos Estadísticos

*Métodos de estimación en regresión de Cox para calcular la
probabilidad de supervivencia académica en estudiantes
universitarios*

Reporte de aplicación para obtener el diploma de Especialista
en
Métodos Estadísticos

Presenta:
Isabel Vega Rebolledo

Director:
Dr. José Juan Muñoz León

Codirectora:
Dra. Cecilia Cruz López

Asesor:
Dr. Ángel Juan Sánchez García

UNIVERSIDAD VERACRUZANA

Facultad de Estadística e Informática

El Comité Académico de la Especialización en Métodos Estadísticos y el director de este trabajo recepcional intitulado: **Métodos de estimación en regresión de Cox para calcular la probabilidad de supervivencia académica en estudiantes universitarios**, autorizan la impresión y la constitución del jurado para la defensa.

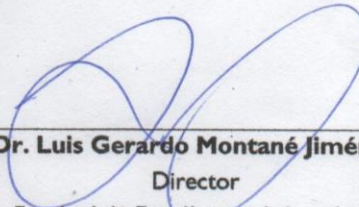
COMITÉ ACADÉMICO



Dra. Cecilia Cruz López

Coordinadora

Especialización en Métodos Estadísticos



Dr. Luis Gerardo Montané Jiménez

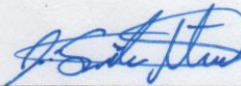
Director

Facultad de Estadística e Informática



**Mtra. María de Lourdes Velasco
Vázquez**

Secretaria



Mtro. José Isaías Siliceo Murrieta

Vocal



Dr. Ángel Fernando Argüello Ortiz

Presidente

GENERACIÓN: 2024-2025

SEDE: Xalapa

TÍTULO: Métodos de estimación en regresión de Cox para calcular la probabilidad de supervivencia académica en estudiantes universitarios

AUTOR: Isabel Vega Rebolledo

DIRECTOR: Dr. José Juan Muñoz León

Codirectora: Dra. Cecilia Cruz López

TIPO DE TRABAJO:

Reporte de aplicación

☒

Trabajo Practico-Educativo

☐

Desarrollo estadístico

☐

Monografía

☐

TIPO DE ESTUDIO:

Exploratorio	☒	Retrospectivo	☐
Descriptivo	☒	Prospectivo	☐
Ex post facto o cuasiexperimental	☐	Transversal	☐
Experimental	☐	Longitudinal	☐

METODOLOGÍA ESTADÍSTICA:

A) Diseño:

Muestreo	☐
Experimento	☐
Estudio observacional	☐

B) Análisis

Exploratorio	☒
Descriptivo básico	☒
Inferencia básica	☒
Métodos multivariados	☐
Regresión	☒
ANOVA y ANCOVA	☐
Control de calidad	☐
Métodos no paramétricos	☒
Modelos especiales	☐
Técnicas avanzadas	☐
Series de tiempo	☐



Universidad Veracruzana

Especialización en Métodos Estadísticos

Facultad de Estadística e Informática

Región Xalapa

11/08/2025

Dra. Cecilia Cruz López
Coordinadora de la Especialización en Métodos Estadísticos

Presente.

Av. Xalapa s/n, Col.
Obrero Campesina,
CP.91020 Xalapa-
Enríquez, Ver., México

Teléfono
22-88-42-17-00

Conm. Ext.
Ext. 14196

Correo electrónico
ceccruz@uv.mx
eme@uv.mx

En mi calidad de director de tesis del C. Isabel Vega Rebolledo, con número de matrícula S24019492 aspirante al grado de Especialista en Métodos Estadísticos, informo a usted que el antes citado ha finalizado el Proyecto de Investigación intitulado Métodos de estimación en regresión de Cox para calcular la probabilidad de supervivencia académica en estudiantes universitarios, cumpliendo con los lineamientos oficiales establecidos por la Coordinación Académica del Posgrado.

De esta manera doy mi visto bueno para que el interesado continúe con el proceso de titulación.

“Lis de Veracruz: Arte, Ciencia, Luz”

Dr. José Juan Muñoz León
Director

25 33 45 45 42 35 59 48 19 71

13 30 27 24 27 31 32 27 12 43

12 3 18 21 15 4 27 21 7 28

51 53 61 31 77 55 29 43 29 45

39 34 44 26 52 28 24 35 15 36

12 19 17 5 25 27 5 8 14 9

34 82 42 73 59 59 43 22 68 52

19 55 27 46 38 32 24 21 41 34

15 27 15 27 21 27 19 1 27 18

43 33 30 73 79 42 73 59 59 43

24 19 25 46 52 27 46 38 32 22

19 14 5 27 27 15 27 21 27 21

68 38 33 33 61 31 51 22 9 74

41 29 19 19 44 26 39 17 8 46

27 9 14 14 17 5 12 5 1 28

A Sergio †:

$$\begin{array}{ccc} 0 & 1 & -1 \\ 2 & -2 & -1 \end{array}$$

Agradecimientos

Al Dr. José Juan, Dr. Ángel Juan y Dr. Ángel Argüello:

Por su amistad, sus consejos, por creer en mí y motivarme a dar un paso más que ayer, sin prisas y sin pausas. Por todo el apoyo brindado a lo largo de este año para mejorar este proyecto. Siempre tendrán un lugar en mi corazón.

A la Dra. Cecilia, Mtra. María de Lourdes y Mtro. Isaías:

Por la orientación y la atención prestada para que este trabajo mejorara.

A la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (Secihti) por la oportunidad brindada para poder seguir con mis estudios.

Índice

Agradecimientos	4
Índice.....	5
Resumen	7
1. Introducción	8
1.1. Planteamiento del problema	9
1.2. Justificación.....	11
1.3. Objetivos.....	11
1.3.1. Objetivo general	11
1.3.2. Objetivos específicos	11
2. Generalidades	12
2.1. Antecedentes.....	12
3. Metodología.....	15
3.1. Población objetivo	16
3.2. Criterios de inclusión.....	16
3.3. Descripción de las variables	16
3.4. Análisis estadístico	18
3.4.1. Análisis exploratorio	18
3.4.2. Modelo de Riesgos Proporcionales de Cox.....	18
3.4.3. Selección de variables	21
3.4.4. Validación de supuestos	21
3.4.5. Evaluación del modelo	22
4. Resultados.....	23
4.1. Análisis exploratorio	23
4.2. Análisis definitivo	28
4.2.1. Modelo de Riesgos Proporcionales de Cox.....	28
4.2.2. Comparación de AIC y validación de supuestos	29
4.2.3. Evaluación del modelo	4
5. Discusión y conclusiones	5
Referencias.....	6

Índice de tablas

TABLA 1. DESCRIPCIÓN DE LAS VARIABLES.....	17
TABLA 2. LICENCIATURA A LA QUE PERTENECEN LOS ESTUDIANTES QUE PERMANECEN Y DESERTARON	23
TABLA 3. MEDIDAS DESCRIPTIVAS PARA EL PUNTAJE CENEVAL	25
TABLA 4. RESULTADOS CON EL MÉTODO DE ESTIMACIÓN DE BRESLOW.....	28
TABLA 5. RESULTADOS CON EL MÉTODO DE ESTIMACIÓN DE EFRON.....	29
TABLA 6. RESULTADOS CON EL MÉTODO DE ESTIMACIÓN VEROSIMILITUD PARCIAL EXACTA.....	29
TABLA 7. COMPARACIÓN DE AIC.....	29
TABLA 8. TEST DE SCHOENFELD.....	30

Índice de Figuras

FIGURA 1. MUNICIPIO DE DOMICILIO DE LOS ESTUDIANTES QUE PERMANECEN Y DESERTARON.....	24
FIGURA 2. GÉNERO DE LOS ESTUDIANTES QUE PERMANECEN Y DESERTARON.....	24
FIGURA 3. PUNTAJES OBTENIDOS EN EL CENEVAL	25
FIGURA 4. PUNTAJES OBTENIDOS EN EL CENEVAL	25
FIGURA 5. ÁREAS QUE CURSARON LOS ESTUDIANTES EN EL NIVEL MEDIO SUPERIOR.....	26
FIGURA 6. ALUMNOS CON BECA EN NIVEL MEDIO SUPERIOR.....	27
FIGURA 7. PROMEDIO GENERAL DE LOS ESTUDIANTES EN LICENCIATURA.....	27
FIGURA 8. AVANCE CREDITICIO DE LOS ESTUDIANTES.....	28
FIGURA 9. TEST DE SCHOENFELD.....	3
FIGURA 10. PRUEBA DE OBSERVACIONES INFLUYENTES	3
FIGURA 11. RESIDUOS DE MARTINGALA.....	3

Resumen

El presente trabajo tiene como objetivo comparar métodos para la estimación de parámetros en la regresión de Cox y con ello identificar los factores que influyen en la supervivencia universitaria de estudiantes de la forma más exacta posible. La base de datos conformada para efectuar dicha comparación contiene 1,338 registros y 22 variables, definiendo como variable respuesta a la supervivencia académica y como covariables a sexo, edad, lugar de procedencia, trayectoria escolar previa, perfil de ingreso, entre otras. Con un AIC de 1401, se identificó que el mejor método para la estimación de los parámetros dentro de la regresión de Cox es Verosimilitud Parcial Exacta, considerando los datos analizados. Por otra parte, promedio general en licenciatura, avance crediticio y puntaje global obtenido en el examen de ingreso se observaron como las variables que mejor explicaron la supervivencia académica universitaria. Mediante el test de Schoenfeld se determinó que tanto el modelo como las variables que lo conformaron cumplen con el supuesto de proporcionalidad de riesgos; además, el modelo obtuvo un c-index de 0.78 y una puntuación de Brier de 0.141, lo que indica que las estimaciones son estadísticamente significativas y confiables.

I. Introducción

La deserción escolar se refiere al proceso gradual mediante el cual los estudiantes se alejan progresivamente de la educación formal, hasta llegar al punto de abandonar completamente la escuela. En el contexto educativo y de acuerdo con Vásquez (2003) este término se utiliza para describir a aquellos estudiantes que dejan de asistir a clases y abandonan sus estudios por diversas razones. Chain (1995), Martínez (2011) y Kuz y Morales (2023), Izasa et al. (2019), entre otros, mostraron que los estudiantes abandonan sus estudios por cuestiones económicas, familiares, sociales, institucionales o por bajo rendimiento académico.

Cabe mencionar que conforme pasa el tiempo, los factores que influyen en la deserción estudiantil pueden cambiar dependiendo de la información disponible que se tenga para analizar. Por ello, en este trabajo se busca analizar este fenómeno mediante una técnica utilizada normalmente en el área de Ciencias de la Salud: el modelo de regresión de Cox. Esto con el fin de identificar los factores del perfil de ingreso y la trayectoria escolar previa de los estudiantes que influyen en la supervivencia académica de los alumnos inscritos entre los años 2019 y 2022 en las licenciaturas en Estadística, Ingeniería de Software, Tecnologías Computacionales y Redes y Servicios de Cómputo, ofertadas en la Facultad de Estadística e Informática (FEI) de la Universidad Veracruzana región Xalapa.

Para ello, se compararán tres métodos para la estimación de los parámetros de la regresión: Efron, Breslow y Verosimilitud Parcial Exacta con los cuales se desea obtener el método que tenga un menor AIC y así poder identificar las variables significativas para el modelo y que influyan en la supervivencia académica de los estudiantes. Con los resultados obtenidos de esta investigación se podrán tomar o mejorar algunas medidas para atacar y disminuir la deserción estudiantil y así poder ayudar a mejorar la eficiencia terminal en las carreras antes mencionadas.

Este trabajo está conformado por cinco secciones. En la sección uno se encuentra el planteamiento del problema, la justificación y los objetivos por los cuales se realizó este proyecto. En la sección dos se describen algunos de los trabajos relacionados al que aquí se presenta. En la sección tres se describe la metodología que se llevó a cabo para dar respuesta a los objetivos planteados. La cuarta sección corresponde a los resultados descriptivos y exploratorios de las variables de la base de datos y los resultados finales en donde se presenta

la regresión de Cox con el mejor método para la estimación de sus parámetros. Finalmente, en la quinta sección se discuten las conclusiones obtenidas.

1.1. Planteamiento del problema

Uno de los tantos temas prioritarios en el sector educativo es la permanencia estudiantil. Aunque no es un tema nuevo, sigue siendo el punto de partida de muchos trabajos y también es uno de los tantos problemas que enfrentan los países. Olarte (2020) indica que la deserción universitaria en los últimos años ha aumentado y en algunos casos la mitad de los alumnos que ingresan al nivel superior, lo abandonan.

La complejidad del proceso de ingresar, permanecer y culminar estudios universitarios ha sido objeto de estudio en los últimos años, ya que numerosos estudiantes enfrentan diversos obstáculos a lo largo de su trayectoria académica. Por ello, en este proyecto se trabajará con alumnos de la Facultad de Estadística e informática de la Universidad Veracruzana región Xalapa donde se ofrecen programas académicos diversificados, distribuidos en cuatro áreas especializadas: 1) Ingeniería de Software 2) Estadística 3) Tecnologías Computacionales y 4) Redes y Servicios de Computo.

Como se ha mencionado, diversos autores han estudiado el tema, como Chain (1995), Velázquez y González (2017), Centurión et al. (2020), entre otros, han identificado que los factores que más influyen son personales, académicos, económicos, culturales, sociales y psicológicos. Cada autor encontró factores diferentes de acuerdo con la información que tenía disponible y analizó. Por ello, en este trabajo se desea abonar a la discusión sobre los factores que influyen en la supervivencia de los estudiantes al momento de cursar una carrera universitaria, intentando dar respuesta a las preguntas de investigación: dentro de la regresión de Cox, ¿Cuál es el mejor método de estimación de parámetros? Y a partir de este método, ¿Cuáles son los factores que influyen para que los estudiantes decidan abandonar sus estudios universitarios?

Al responder estas preguntas, lo que se pretende en primera instancia es saber si factores que aún no se han sido mencionados en la literatura, como el promedio general en licenciatura y el avance crediticio, influyen en la probabilidad de supervivencia del estudiante y en qué manera lo hacen, aplicando el modelo de riesgos proporcionales de Cox. Este modelo de regresión permite reportar en qué medida ciertos factores condicionan la probabilidad de

supervivencia en un determinado contexto, en este caso se busca comparar diferentes métodos para calcular los estimadores que dan lugar a dicha forma de regresión. Concretamente, se compararán los métodos de estimación de Breslow, Efron y Verosimilitud Parcial Exacta, este último propuesto por Cox. A partir de esta comparación se pretende reportar cuál método de estimación explica mejor el fenómeno que en esta investigación se abordará, esto mediante su AIC. Dicha comparación tiene lugar debido a que, la mayor parte de los softwares estadísticos aplican automáticamente el método de Efron, ya que es computacionalmente más intensivo y esto hace suponer que es más eficiente que el método de Breslow y Verosimilitud Parcial Exacta. Adicionalmente, en la literatura se han encontrado reportes de investigación como el de Ramírez et al. (2017) quienes realizaron la comparación de estos métodos e indican que no se encontraron diferencias significativas entre estos métodos.

1.2. Justificación

Algunos trabajos analizados como el de Chain (1995) indica que para predecir el éxito o fracaso en la trayectoria escolar de un estudiante de nivel superior es indispensable fijarse en factores personales, culturales, familiares, socioculturales, académicos y normativos. Por otro lado, Velázquez y González (2017) indican que los factores que están asociados a la permanencia en estudiantes de educación superior son motivacionales, académicos, sociales, familiares y económicos. En este sentido, se pretende abonar a la discusión sobre si los datos de la trayectoria escolar, el perfil de ingreso y datos de los estudiantes al momento de cursar carrera, como es el promedio general de licenciatura y el avance crediticio, influyen en la supervivencia académica de los estudiantes inscritos en primera lista durante 2019 - 2022 en las carreras de Ingeniería de Software, Estadística, Tecnologías Computacionales y Redes y Servicios de Cómputo. Esto debido a que se ha observado que, en el periodo de estudio, disminuyó la eficiencia terminal en las carreras antes mencionadas. Es por ello que, se pretende identificar el mejor método de estimación para los parámetros de la regresión de Cox para así poder identificar los factores que influyen en la supervivencia de los estudiantes. Con esta información se pretende ayudar tanto a la institución como los docentes ya que así se podrán identificar a los estudiantes que tengan riesgo de abandono escolar.

1.3. Objetivos

1.3.1. Objetivo general

Comparar los métodos para la estimación de los parámetros en la regresión de Cox aplicándolos sobre los datos de permanencia y deserción en estudiantes universitarios.

1.3.2. Objetivos específicos

Aplicar métodos de estimación de Breslow, Efron y Verosimilitud Parcial Exacta e identificar cuál es el mejor para el modelo de Cox a partir de su AIC.

Identificar los factores que influyan en la probabilidad de supervivencia de los estudiantes universitarios.

2. Generalidades

A continuación, se presentan algunos de los trabajos relacionados.

2.1. Antecedentes

Ramírez et al. (2017) compararon los métodos de estimación de Breslow, Efron y Exacto en Regresión de Cox, aplicándolos a datos simulados y obteniendo los intervalos de confianza mediante remuestreo Bootstrap, Jackknife y Anderson. En donde concluyen que no existen diferencias significativas entre los tres estimadores comparados.

Vergel et al. (2016) utilizaron análisis de supervivencia para identificar los factores que influyen en la deserción universitaria. Trabajaron con estudiantes de diferentes carreras de la Universidad Francisco Paula Santander en Cúcuta, Colombia. Inicialmente aplicaron el análisis de supervivencia de Kaplan-Meier para obtener la función de supervivencia y posteriormente el modelo de riesgos proporcionales de Cox en donde obtuvieron que las variables que influyen en el tiempo de permanencia de los estudiantes son factores individuales y socioeconómicos, académicos y factores institucionales.

Vergara et al. (2016) identificaron los factores que influyen en la deserción de estudiantes de pedagogía de una universidad de Chile. La base consta de 531 registros de estudiantes de la generación 2009–2013, las variables que analizaron eran variables individuales, académicas, socioeconómicas e institucionales. Dichos autores indican que las variables que influyen en la deserción de los estudiantes son individuales como sexo y región de procedencia y académicas: como promedio en nivel medio superior, lugar en la lista de seleccionados, que el estudiante provenga de un establecimiento secundario científico-humanista, número de materias inscritas y promedio.

Daza (2016) generó un modelo basado en árboles de decisión para predecir la deserción estudiantil en universitarios provenientes de escuelas particulares. En este trabajo, se incluyen variables como identificación del alumno, rendimiento académico del colegio, edad, sexo, entre otras teniendo un total de 30 variables. Indica que las variables que se deben tomar en cuenta son código de Universidad, rendimiento académico de la institución, rendimiento académico del colegio, edad, sexo, ciclo, cantidad de créditos aprobados, créditos desaprobados, entre otros.

Puma (2018) utilizó el análisis de supervivencia de Kaplan-Meier, la prueba de Long-Rank y el modelo de Cox para estimar la probabilidad de supervivencia a la deserción en

estudiantes de la Facultad de Ciencias Naturales y Formales de la Universidad Nacional de San Agustín. Contaba con datos como escuela, género, colegio, edad de ingreso a universidad, tiempo que el alumno tardó en ingresar a la universidad y lugar de estudios. Dicho autor concluye que los factores que influyen en la deserción estudiantil de la facultad son el colegio de procedencia, edad del alumno y el lugar de estudios. Indica que los alumnos que tuvieron mayor deserción son aquellos que provienen de un colegio nacional, mayores de 25 años y que estudian en ciudades.

Isaza et al. (2019) utilizaron el modelo de Cox para analizar si factores individuales, académicos y socioeconómicos interfieren en el tiempo de permanencia y deserción de estudiantes de una universidad de Colombia. Para ello utilizaron datos de alumnos inscritos entre 2008 y 2014 de todas las carreras que oferta esta universidad. Indican que los factores que interfieren con la permanencia y deserción estudiantil son académicos como la carrera y número de periodos académicos cursados y variables socioeconómicas como son el género y el ingreso familiar.

Centurión et al. (2020) generaron un modelo basado en una red neuronal para la predicción del abandono escolar en estudiantes de ingeniería. Este modelo se generó a partir de datos del perfil socioeconómico, académico y personal. El modelo tuvo una exactitud del 75%. Sin embargo, los autores recalcan que la información analizada no resultó significativa para poder predecir el abandono escolar.

Álvarez et al. (2024) aplicaron análisis de supervivencia para determinar qué variables se asocian al tiempo de supervivencia de estudiantes universitarios. Para ello, contaban con 421 registros de estudiantes de la Facultad de Ciencias Económicas de la Universidad Hispanoamericana. Como variable dependiente, tomaron el tiempo de supervivencia de los estudiantes y como variables independientes variables como género, sede del estudiante, edad, trabajo, carrera, beca, colegio de procedencia, nacionalidad y provincia. Posteriormente, aplicaron la prueba log-rank para probar que no existen diferencias entre las poblaciones. En donde, con un nivel de significancia del 0.05, indican que no encontraron diferencias significativas entre las curvas de supervivencia y las variables explicativas. Finalmente, mediante el modelo de Cox, indican que las variables que influyen en la supervivencia de los estudiantes son Sede, Colegio de procedencia y Provincia.

Lazo et al. (2024) aplicaron análisis de supervivencia de Kaplan-Meier y el Modelo de riesgos proporcionales de Cox para determinar el tiempo que transcurre para que el estudiante deserte en el nivel superior. Analizaron una base de datos de 4424 registros y 36 variables en las que se incluían datos académicos, demográficos, factores socioeconómicos y rendimiento académico de los estudiantes durante los dos primeros semestres. Indican que las variables que influyen en la deserción son edad, carrera y materias aprobadas en primer y segundo semestre.

Osorio (2024) aplicó análisis de supervivencia y aprendizaje automático para analizar e identificar los factores más relevantes para predecir la deserción de estudiantes universitarios. Dicho autor, analizó variables sociodemográficas en las que se incluyen características personales como género, edad, entre otras; académicas como promedio, calificaciones, asistencia a clases, entre otras y sociocognitivas como la motivación, satisfacción compromiso, entre otras. En donde obtuvo que las variables que debemos observar son las académicas, personales, sociodemográficas, entre otras. Destaca que, para este estudio, se deben observar más las variables sociodemográficas ya que estas juegan un papel crucial en la predicción de la deserción estudiantil.

Martínez et al. (2024) aplicaron el análisis de supervivencia de Kaplan-Meier y el modelo de Cox para analizar las trayectorias académicas de la licenciatura en medicina de la UNAM para identificar factores que interfieran en la regularidad de los estudiantes. Utilizaron variables como género, bachillerato de procedencia, regularidad y eficiencia terminal, entre otras. Concluyen que es indispensable mirar el género y el bachillerato de procedencia ya que estas variables interfieren en la eficiencia terminal.

3. Metodología

Los datos utilizados para llevar a cabo este trabajo fueron extraídos del Sistema de Planeación Académica (PLANEA) y de la plataforma scOpi Web¹, la cual es dirigida por el Centro de Investigaciones e Innovación en Educación Superior (CIIES)² y la Dirección General de Administración Escolar (DGAE)³, ambos pertenecientes a la Universidad Veracruzana. En estas plataformas se pueden consultar datos del perfil de ingreso y trayectoria escolar previa de los aspirantes que presentaron el Examen Nacional de Ingreso a la Educación Superior (EXANI-II) el cual es una herramienta para la evaluación de aptitudes y conocimientos de los aspirantes a la educación superior. Esta prueba la aplican aquellas instituciones que hayan contratado los servicios del CENEVAL. De igual forma, se pueden consultar datos de los estudiantes al momento de cursar la carrera, como promedio general de licenciatura, promedio no ponderado, avance crediticio, calificaciones, estatus del estudiante, entre otras.

La base de datos fue capturada en Excel, cuenta con 1,338 registros de estudiantes inscritos durante el periodo 2019 a 2022 en las licenciaturas en Ingeniería de Software, Estadística, Tecnologías Computacionales y Redes y Servicios de Cómputo. Contiene 22 variables de las cuales 7 son de carácter cuantitativo y 15 de carácter cualitativo, que a su vez se dividen en 4 apartados: 1) datos generales del alumno como lugar de nacimiento, edad, sexo, entre otras; 2) Trayectoria escolar previa como promedio de bachillerato, área cursada en bachillerato; 3) Perfil de ingreso como puntaje global del examen de ingreso (EXANI-II); y 4) datos del alumno al momento de cursar la licenciatura como avance crediticio y promedio general. Adicionalmente, para realizar los análisis pertinentes se añadió una variable tiempo en donde se especifica el inicio o fecha en que el alumno ingresó a la

¹ <https://scopi.aexiuv.com/>

² <https://www.uv.mx/ciies/>

³ <https://www.uv.mx/dgae/>

licenciatura, fin o fecha en que el alumno desertó o permanece en el programa educativo y la variable supervivencia la cual indica si el alumno desertó, sigue activo o egresó (véase en **Tabla 1**)

3.1. Población objetivo

Alumnos que ingresaron a las licenciaturas en Estadística, Ingeniería de Software, Tecnologías Computacionales y Redes y Servicios de Cómputo región Xalapa durante el periodo comprendido entre 2019 y 2022.

3.2. Criterios de inclusión

Alumnos con derecho a inscripción en primera lista (1A), Vacante (2A), primer corrimiento (1C) y tercera vuelta en vacante (3B) que se matricularon durante el periodo de tiempo antes mencionado.

3.3. Descripción de las variables

En la tabla 1 se muestran las variables que contiene la base de datos, así como las diferentes respuestas que dieron los alumnos.

Tabla 1. Descripción de las variables

Variable	Respuesta	Tipo de variable	Escala
Año	De 2019 a 2022	Cuantitativa	Discreta
Carrera	Estadística, Ingeniería de Software, Tecnologías computacionales y Redes y servicios de computo	Cualitativa	Nominal
Municipio de domicilio	Xalapa / Otro	Cualitativa	Nominal
Sexo	Masculino / Femenino	Cualitativa	Nominal
Puntaje CENEVAL	22 pts hasta 97 pts	Cuantitativa	Continua
Turno	Matutino / Vespertino / Nocturno / Mixto / Discontinuo / No especificado / Otro	Cualitativa	Nominal
Edad	16 a 49 años	Cuantitativa	Continua
Área en bachillerato	Artes / Biológicas / Ciencias de la Salud / Económico Administrativas, Humanidades / Técnica / No especificado	Cualitativa	Nominal
Promedio de bachillerato	0 a 10	Cuantitativa	Continua
Régimen	Público / Privado / Otro	Cualitativa	Nominal
Modalidad	Bachillerato general / Bachillerato tecnológico / Profesional técnico / Telebachillerato	Cualitativa	Nominal
Nivel máximo de estudios a alcanzar	Licenciatura / Posgrado / Técnico Superior Universitario	Cualitativa	Ordinal
Trabaja	Sí / No	Cualitativa	Nominal
Beca en bachillerato	Sí / no	Cualitativa	Nominal
Escolaridad de la madre	No especificado / No estudió / Básica / Media Superior / Superior	Cualitativa	Ordinal
Escolaridad del padre	No especificado / No estudió / Básica / Media Superior / Superior	Cualitativa	Ordinal
Promedio general	1 a 10 pts.	Cuantitativa	Continua
Promedio no ponderado	0 a 10	Cuantitativa	Continua
Avance crediticio	0 a 110 créditos	Cuantitativa	Continua
Supervivencia	0 = deserta / 1 = permanece o egresó	Cualitativa	Nominal
Fecha de inicio	agosto de 2019 hasta agosto de 2022	Cualitativa	Nominal
Fecha de fin	enero de 2020 hasta enero de 2024	Cualitativa	Nominal

3.4. Análisis estadístico

3.4.1. Análisis exploratorio

Como primer acercamiento, se realizó un análisis exploratorio de los datos con el fin de tener un panorama general sobre su comportamiento. Este primer acercamiento consistió en generar tablas de frecuencia y algunos gráficos para poder observar el comportamiento de las variables estudiadas.

3.4.2. Modelo de Riesgos Proporcionales de Cox

El análisis de supervivencia es un método estadístico utilizado en el ámbito médico, donde la variable de estudio es el tiempo que transcurre antes de que ocurra un evento. El tiempo de inicio y el fin del evento es el foco de este análisis. Cabe mencionar que el tiempo se puede medir en días, semanas o incluso meses.

El modelo de Cox fue propuesto por Cox en 1972, es un modelo semiparamétrico el cual incluye una parte paramétrica y otra no paramétrica. En mucha de la literatura suele ser llamado modelo de riesgos proporcionales de Cox ya que, el riesgo para dos individuos cualesquiera es constante e independiente al tiempo de supervivencia de estos. A diferencia de Kaplan-Meier, Cox permite añadir covariables o variables explicativas, ya que el objetivo es modelar la relación entre las variables explicativas y el tiempo de supervivencia de los individuos, además permite identificar las variables que influyen para que suceda el evento de interés. Como en todo modelo de regresión, se deben estimar los coeficientes (β), sin embargo, al ser un modelo semiparamétrico no se pueden estimar los coeficientes mediante el método de Máxima Verosimilitud, por lo que Cox propuso hacer la estimación de los coeficientes mediante el método de Verosimilitud Parcial Exacta para así no solo obtener la probabilidad de supervivencia de aquellos individuos que mueren, sino que también obtener la probabilidad de supervivencia de aquellos individuos censurados (individuos que no han pasado por el evento que se desea estudiar durante el tiempo de observación).

El modelo está dado por la siguiente expresión:

$$h(t|x) = h_0(t) \cdot \exp(x'\beta)$$

donde:

$h(t|x)$ es el tiempo (t) dado un conjunto de variables explicativas (x)

$h_0(t)$ es el riesgo basal

x' es el vector conformado por las variables explicativas

β coeficientes de las variables explicativas

3.4.2.1. Métodos de estimación en regresión de Cox

Al hablar de supervivencia, es muy probable que tengamos el problema de empates en los tiempos de supervivencia. Es recomendable aplicar métodos de estimación distintos al de Máxima Verosimilitud que es el que se utiliza cuando no hay empates en los tiempos de supervivencia de los individuos. Algunos de los métodos de estimación son los siguientes:

3.4.2.2. Método de estimación de Efron

El método de Efron es también otro estimador para el modelo de Cox. Este método utiliza una aproximación delicada la cual consiste en ir asignando pesos fraccionados a cada individuo en el tiempo de empate, es decir, el tiempo de supervivencia lo dividirá en intervalos y a cada intervalo le asignará un peso, esto con el fin de obtener estimaciones de los coeficientes más precisas.

Fue propuesto en 1977, es una alternativa al método de Breslow y da aproximaciones más precisas (Ramírez, Regino y Guerrero, 2017). Asume que, cuando hay empates en los tiempos de supervivencia, se debe distribuir el riesgo de manera proporcional entre los individuos que empatan.

Los parámetros de esta estimación están dados por la expresión:

$$L_E(\beta) = \prod_{i=1}^D \frac{\exp(x_i\beta)}{\sum_{k \in R(t_i)} \exp(x_k\beta) - \frac{d_j}{m_i} \sum_{k \in U_{(i)}} \exp(x_k\beta)}$$

x_i variables explicativas de aquellos individuos que experimentan el evento de interés en un tiempo dado.

d_j numero de eventos que ocurren simultáneamente

$\frac{d_j}{m} \sum_{k \in U_{(i)}} \exp(x_k\beta)$ distribución del riesgo entre los individuos

3.4.2.3. *Método de estimación de Breslow*

Propuesto por Breslow en 1974. Este método se aproxima a la Verosimilitud Parcial de forma sencilla y se considera con menor complejidad computacional; este método trata a los empates en los tiempos de supervivencia como si los eventos sucedieran uno tras otro.

La fórmula está dada por:

$$L_B(\beta) = \prod_{i=1}^D \frac{\exp(x_i\beta)}{\sum_{j \in R(t_i)} \exp(x_j\beta)}$$

D es el número de empates

$\exp(x_i\beta)$ exponencial de las covariables con sus respectivos coeficientes

$R(t_i)$ aquellos individuos que aún no sufren el evento de interés pero siguen en riesgo de sufrirlo

$\sum_{j \in R(t_i)} \exp(x_j\beta)$ es la suma de riesgos de los individuos en el tiempo t_i y hasta que ocurre el evento de interés.

3.4.2.4. *Método por Verosimilitud Parcial Exacta*

Se considera el más intensivo de los métodos ya que, simultáneamente, calcula la probabilidad de que el evento ocurra en el mismo tiempo para un grupo de individuos considerando todas las posibles permutaciones entre eventos empatados.

$$L_{VPE}(\beta) = \prod_{i=1}^D \frac{\exp(x_i\beta)}{\sum_{u \in U_{(i)}} \exp(x_u\beta)}$$

Donde:

D número total de fallas

x_i vector de las variables explicativas de aquellos individuos que experimentan el evento de interés en el tiempo

$U_{(i)}$ conjunto de individuos que no fallan ni son censurados

x_j individuos que no experimentan el evento de interés en el tiempo establecido

3.4.3. Selección de variables

Para seleccionar las variables que aportan a cada uno de los modelos generados, se utilizó backward regression (regresión hacia atrás), esta técnica se utiliza para identificar aquellas variables que son importantes para el modelo. Este tipo de regresión comienza con un modelo con todas las variables predictoras, llamado modelo saturado, el cual está dado por la expresión:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon$$

Donde:

$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$: Todas nuestras variables predictoras
 ε : es el error

Posteriormente el algoritmo irá agregando o eliminando variables del modelo de manera automática, ya que utiliza un criterio de selección. Al finalizar esto, arroja el mejor modelo con un subconjunto de variables que mejoren el rendimiento de este.

El criterio de selección más utilizado es el p-valor, el cual indica la probabilidad de que el coeficiente de una variable sea cero. En este caso, el criterio de selección fue el Criterio de Información de Akaike (AIC) el cual toma en cuenta el rendimiento del modelo y la complejidad de él. A menor AIC mejor es el modelo.

3.4.4. Validación de supuestos

Como en todo modelo de regresión, se deben cumplir tres supuestos importantes: proporcionalidad de riesgos, observaciones influyentes y no linealidad.

Para verificar el supuesto de proporcionalidad de riesgos se aplicó la prueba de Schoenfeld. De acuerdo con Keymolen et al. (2017) la prueba de Schoenfeld consiste en estimar la diferencia entre el valor predicho y el esperado. Esto se hace para ver si, transcurrido el tiempo el efecto de las covariables cambia.

Para identificar que las covariables cumplan con el supuesto de proporcionalidad de riesgos se debe observar el valor de p y mediante las hipótesis:

H_0 : se cumple el supuesto de proporcionalidad de riesgos

H_1 : no se cumple el supuesto de proporcionalidad de riesgos

Posteriormente se realizó la prueba de observaciones influyentes, en la cual se obtuvo el gráfico de los cambios en los coeficientes del modelo el cual ayuda a identificar observaciones que puedan distorsionar los resultados. Finalmente, para comprobar si la relación entre el riesgo y cada una de las variables que salieron significativas es lineal o requieren transformación se realizó el gráfico de residuos de Martingala.

3.4.5. Evaluación del modelo

Para evaluar la capacidad predictiva en regresión de Cox se aplicaron dos métricas:

- 1) Índice de concordancia (C-index): También conocido como índice C de Harell, es una métrica de bondad de ajuste utilizada para evaluar la capacidad discriminativa del modelo, es decir, qué tan bien distingue el modelo a los sujetos con mayor y menor riesgo a sufrir el evento de interés. Algunos autores como Allende (2019) lo definen como la proporción de pares concordantes dividida entre el número de pares no concordantes o pares de evaluación posibles.

Hartman et al. (2023) indican que un valor de 0.5 corresponde al rendimiento de un clasificador aleatorio. Lo óptimo son valores superiores a 0.7 ya que esto indica que el modelo discrimina adecuadamente entre los perfiles de riesgo.

- 2) Puntuación de Brier (Brier Score): Se puede utilizar en eventos binarios o categóricos siempre y cuando, las categorías puedan tener la estructura de una variable binaria. Cuando la puntuación de Brier es 0, significa que hay una precisión total o perfecta. Mientras que cuando la puntuación es 1, quiere decir que el pronóstico es inexacto. Lo ideal son las puntuaciones cercanas a 0, ya que éstas indican mejores pronósticos. La fórmula para calcular la puntuación de Brier es:

$$BS = \frac{1}{N} \sum_{t=1}^N (f_t - o_t)^2$$

f_t : probabilidad prevista

o_t : resultado (1 sucede o 0 no sucede el evento)

Para obtener la puntuación de Brier se generaron y compararon tres modelos: modelo de referencia que es el modelo “ingenuo” que asigna una probabilidad fija para todos los individuos, es decir, todos tendrán la misma probabilidad de sobrevivir; modelo nulo el cual

no incluye a las variables significativas y finalmente el modelo de Cox generado originalmente. Esto con el fin de verificar el modelo obtenido tenga un buen desempeño predictivo.

4. Resultados

4.1. Análisis exploratorio

A continuación, se muestran los resultados del análisis exploratorio de los datos.

Tabla 2. Licenciatura a la que pertenecen los estudiantes que permanecen y desertaron

Licenciatura	Porcentaje de egreso
Ingeniería de Software	31%
Estadística	45%
Redes y Servicios de Cómputo	56%
Tecnologías Computacionales	45%
General	44%

La tabla 2 muestra las licenciaturas y el porcentaje de egreso en cada una de ellas, en donde se observa que en la licenciatura en Ingeniería de Software los estudiantes fueron propensos a desertar. Y en general, de estas 4 licenciaturas sólo el 44% de los estudiantes egresa.

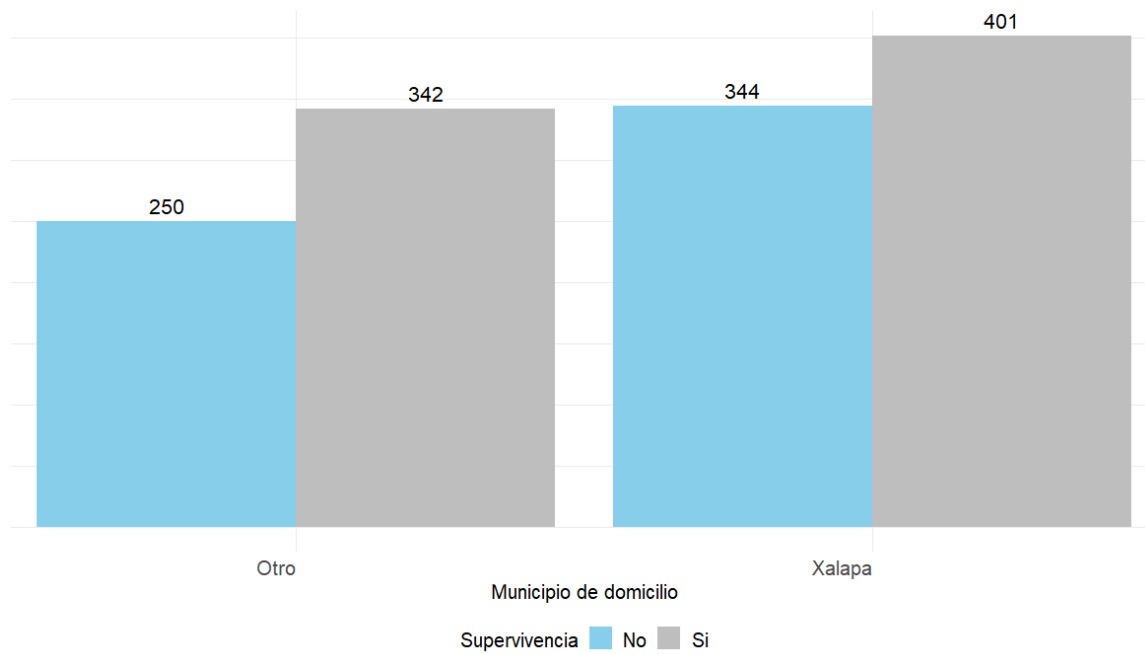


Figura 1. Municipio de domicilio de los estudiantes que permanecen y desertaron.

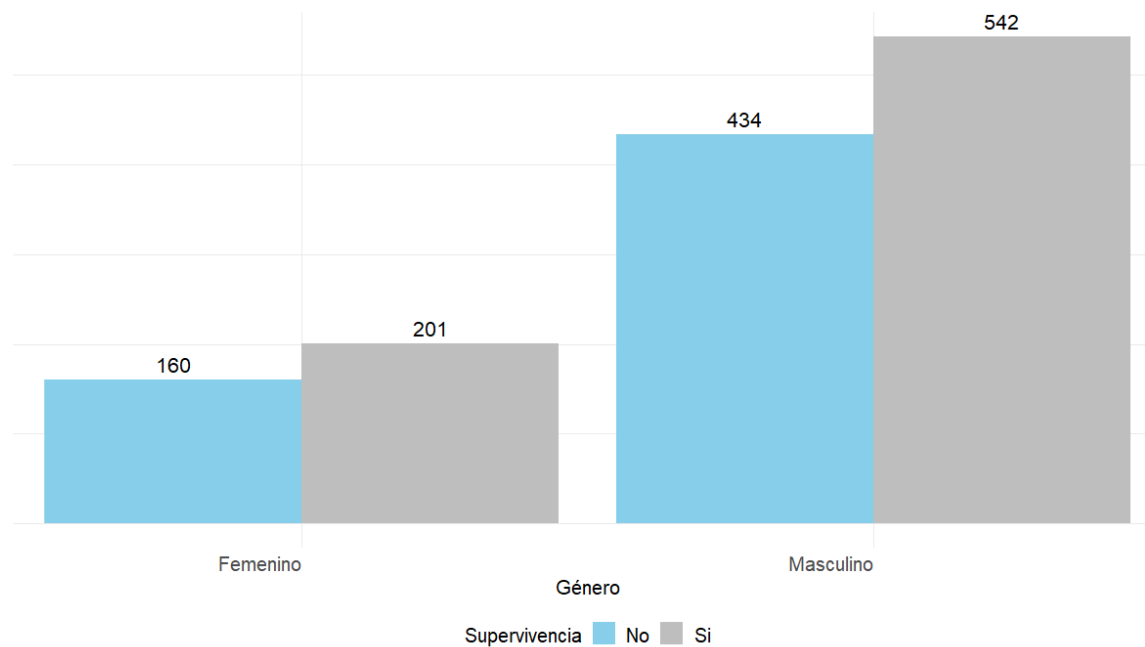


Figura 2. Género de los estudiantes que permanecen y desertaron.

La figura 1 y 2 permiten explorar el comportamiento de la supervivencia en estudiantes residentes y foráneos, en donde se observa que los estudiantes que permanecen (401 alumnos) y también los que son más propensos a desertar (344 alumnos) residen en Xalapa.

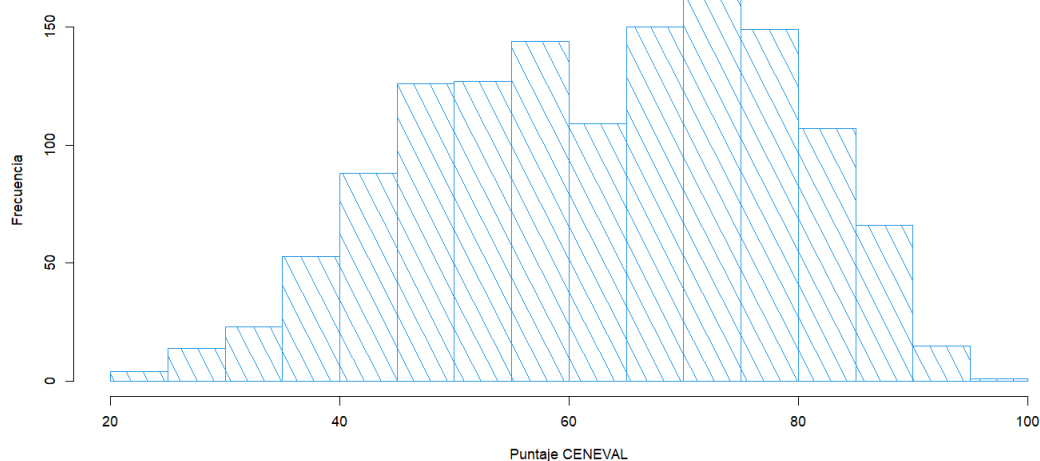


Figura 3. Puntajes obtenidos en el CENEVAL

Tabla 3. Medidas descriptivas para el puntaje CENEVAL

Valor Mínimo	1 ^{er} cuartil	Mediana	Media	3 ^{er} cuartil	Valor Máximo
22	51	64.36	63.28	73.31	97

La figura 3 muestra cómo se distribuyen los puntajes obtenidos en el CENEVAL por los estudiantes. Como se observa, la mayoría obtuvieron entre 65 y 80 puntos. Esta figura indica una leve asimetría a la izquierda ya que la media (63.28) es menor que la mediana (64.36), esto también se puede corroborar al observar el valor de los cuartiles con la figura 4 en donde se observa que la distancia mayor es entre la mediana y el primer cuartil, o sea hacia el lado izquierdo.

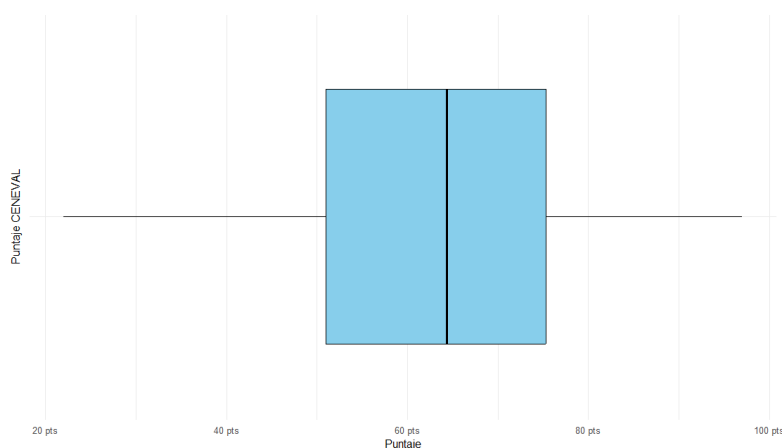


Figura 4. Puntajes obtenidos en el CENEVAL.

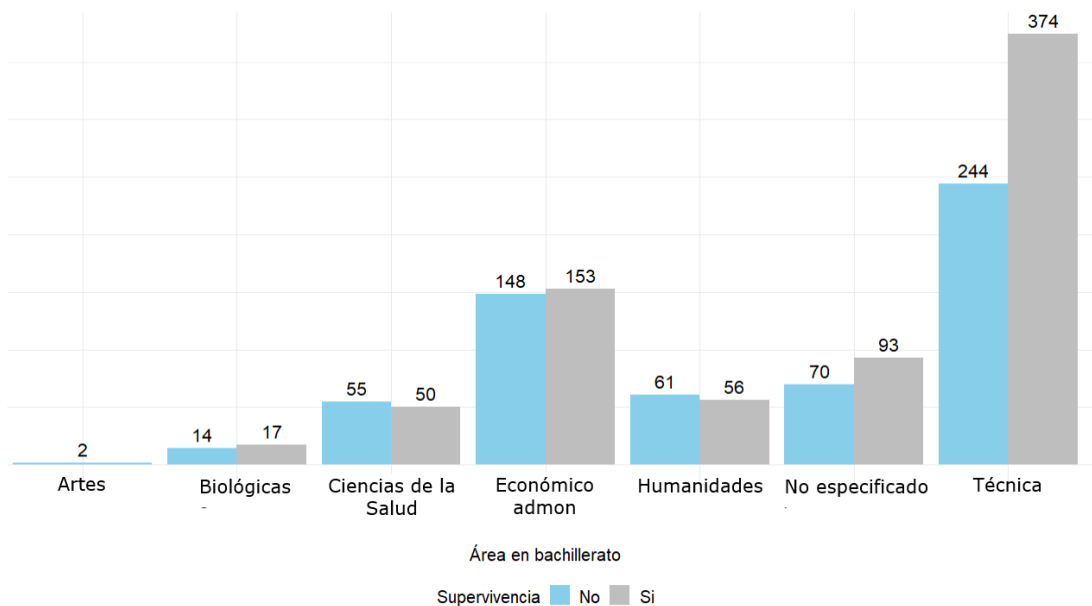


Figura 5. Áreas que cursaron los estudiantes en el nivel medio superior.

La figura 5 corresponde al área cursada en el nivel medio superior, siendo los estudiantes provenientes del área técnica los que predominan y permanecen en el nivel superior, seguido de los estudiantes del área económico-administrativa.

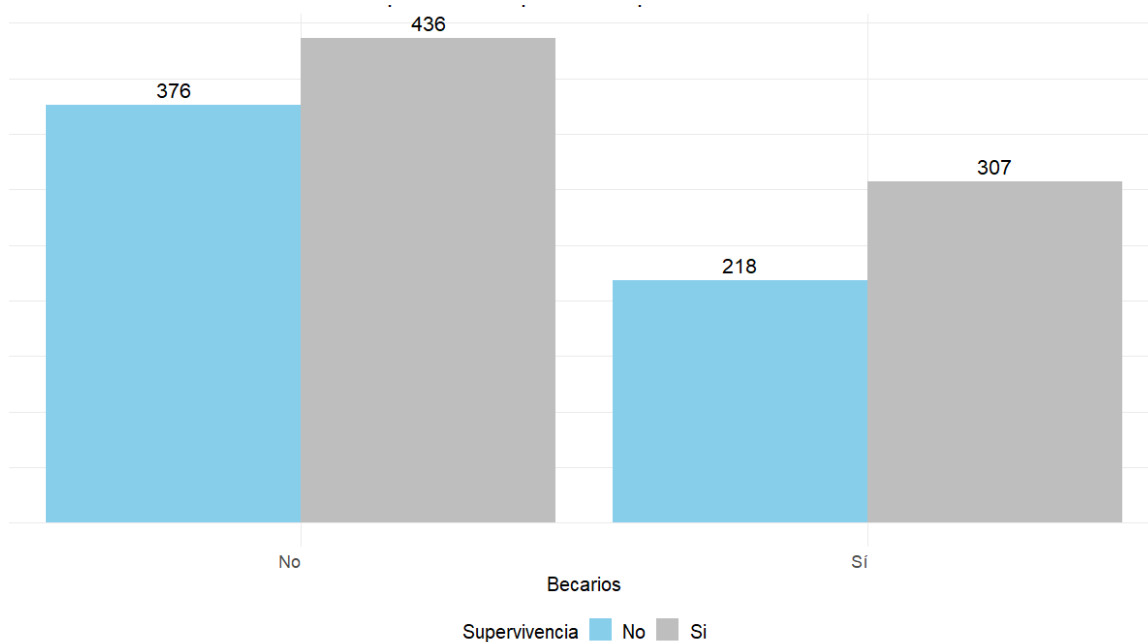


Figura 6. Alumnos con beca en nivel medio superior.

La figura 6 representa a los alumnos que tuvieron beca por necesidad económica en el nivel medio superior. Cabe señalar que aquellos que no tuvieron este tipo de beca permanecen en el nivel superior.

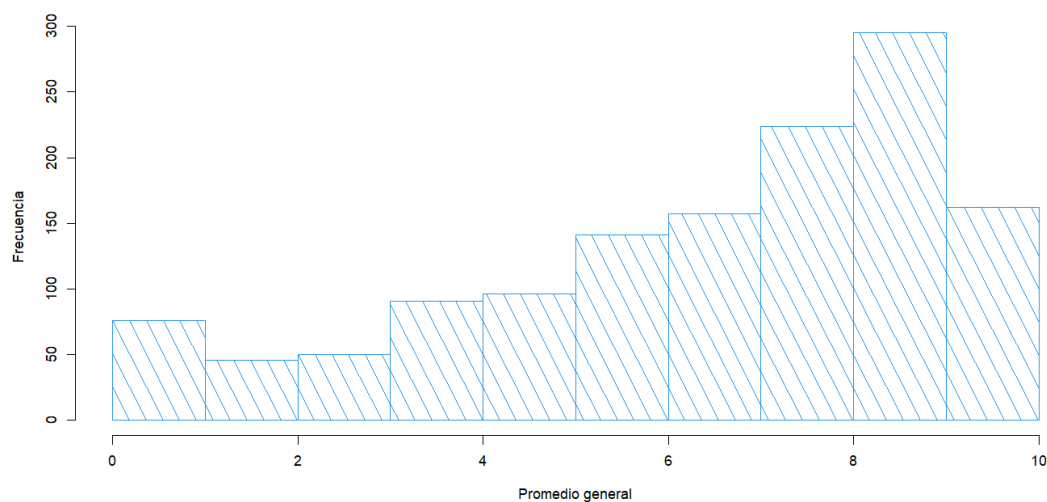


Figura 7. Promedio general de los estudiantes en licenciatura.

La figura 7 representa el promedio general de los estudiantes en el nivel superior, en donde se observa que la mayoría de los estudiantes tiene un promedio general entre 7 y 9.

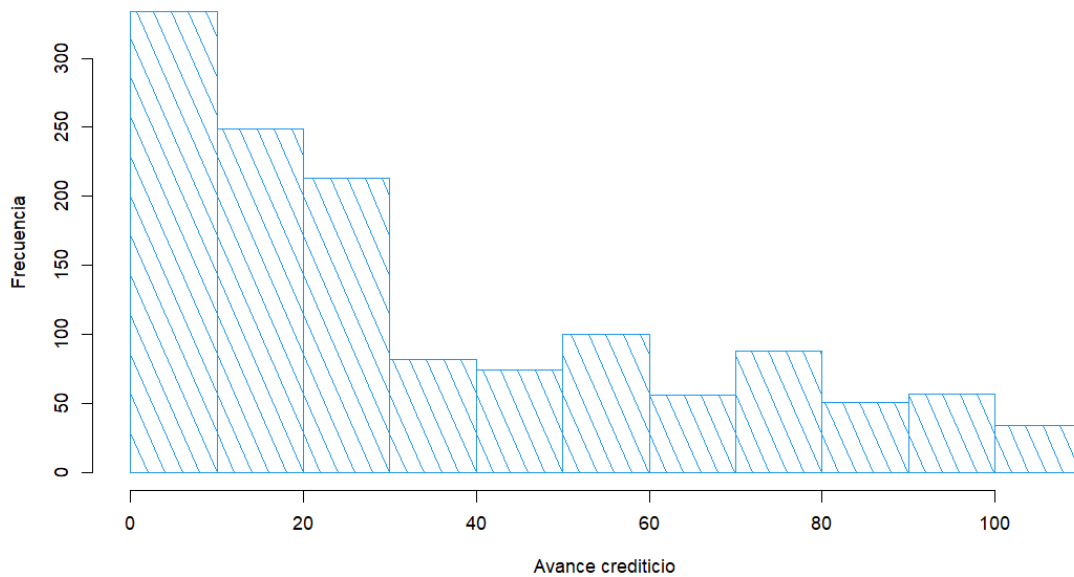


Figura 8. Avance crediticio de los estudiantes.

Finalmente, la figura 8 corresponde al avance crediticio de los estudiantes, en donde se observa que la mayoría de ellos tienen entre el 0 y 30% de avance crediticio.

4.2. Análisis definitivo

4.2.1. Modelo de Riesgos Proporcionales de Cox

Para generar el modelo de regresión de Cox con cada uno de los métodos antes mencionados, la base de datos se dividió al 80-20, donde 80% de los datos se utilizó para entrenar al modelo y el 20% restante para pruebas en donde se obtuvieron los resultados que a continuación se presentan.

Tabla 4. Resultados con el método de estimación de Breslow.

	<i>β</i>	<i>Error estándar</i>	<i>z</i>	<i>Sig.</i>
Año 2020	0.187334	0.036496	5.133	< 0.000 ***
Año 2021	0.048339	0.003528	13.702	< 0.000 ***
Año 2022	0.888156	0.162820	5.455	< 0.000 ***
Promedio general	0.3575	0.0443	8.062	< 0.000 ***
Avance crediticio	-0.0295	0.0032	-9.208	< 0.000 ***
Puntaje CENEVAL	-0.009371	0.004079	-2.297	0.0216 *

Signif. codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Tabla 5. Resultados con el método de estimación de Efron.

	<i>β</i>	Error estándar	<i>z</i>	Sig.
Año 2020	0.925377	0.175311	4.110	0.00003 ***
Año 2021	2.144430	0.222894	12.680	< 0.000 ***
Año 2022	3.488310	0.284531	5.278	< 0.000 ***
Promedio general	0.164858	0.040107	9.621	< 0.000 ***
Avance crediticio	0.049111	0.003873	12.260	< 0.000 ***

Signif. codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Tabla 6. Resultados con el método de estimación Verosimilitud Parcial Exacta.

	<i>β</i>	Error estándar	<i>z</i>	Sig.
Puntaje CENEVAL	-0.017335	0.00502	-3.094	0.00197**
Promedio general	1.6491119	0.15860	10.398	0.01646 **
Avance crediticio	-0.110531	0.01025	-10.774	< 0.000 ***

Signif. codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Como se observa en las tablas 4 a 6, las variables en común para los modelos generados con los métodos de estimación fueron promedio general en licenciatura, avance crediticio y puntaje obtenido en el CENEVAL, esta última apareciendo en el método de estimación de Breslow y Verosimilitud Parcial Exacta. Adicionalmente, en el método de Efron y Breslow indican que también los años de 2020 a 2022 son relevantes para el modelo.

4.2.2. Comparación de AIC y validación de supuestos

Tabla 7. Comparación de AIC.

Método de estimación	AIC
Breslow	3938.834
Efron	2541.844
Verosimilitud Parcial Exacta	1401.126

Para identificar el mejor método para la estimación de los parámetros de la regresión, se obtuvo el AIC. En este caso, el modelo con el método de Verosimilitud Parcial Exacta es el que arrojó un menor AIC:1401.126, como se observa en la tabla 7. Una vez seleccionado el mejor método de estimación para el cálculo de los parámetros del modelo de Cox, se realizó el test de residuos de Schoenfeld, con el cual se busca que las variables cumplan el supuesto de proporcionalidad de riesgos.

Tabla 8. Test de Schoenfeld.

	χ^2	<i>Grados de libertad</i>	<i>Valor p</i>
Puntaje CENEVAL	0.193	1	0.66
Promedio general	0.253	1	0.62
Avance crediticio	1.168	1	0.28
GLOBAL	1.273	3	0.74

Como se observa en la tabla 8, los valores p en todas las variables son mayores al nivel de significancia (0.05), lo cual indica que las variables del modelo cumplen el supuesto de proporcionalidad de riesgos. También se obtiene el valor p del modelo global (0.74) el cual también cumple con el supuesto de proporcionalidad de riesgos, esto se visualiza en la figura 9.

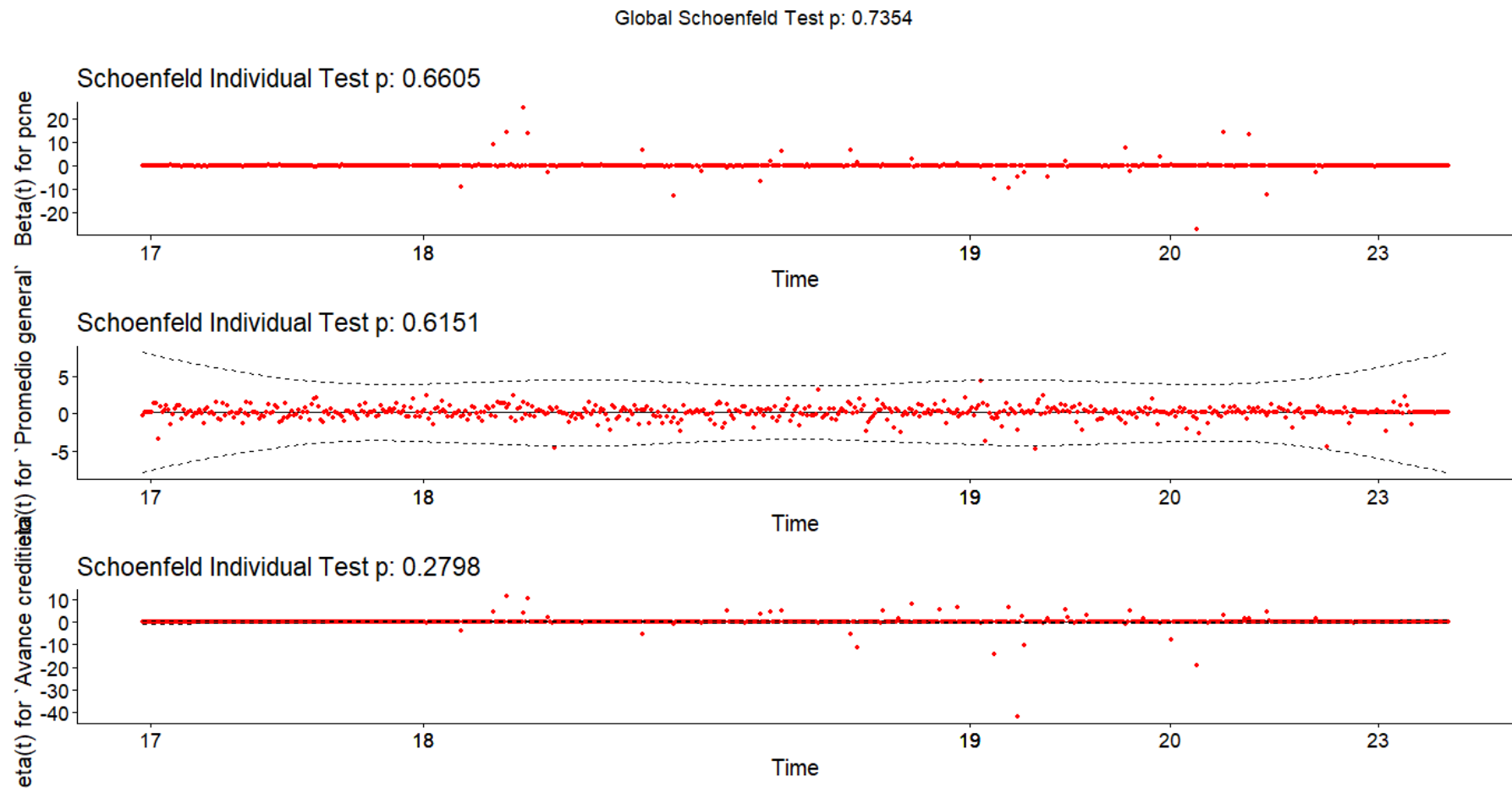


Figura 9. Test de Schoenfeld.

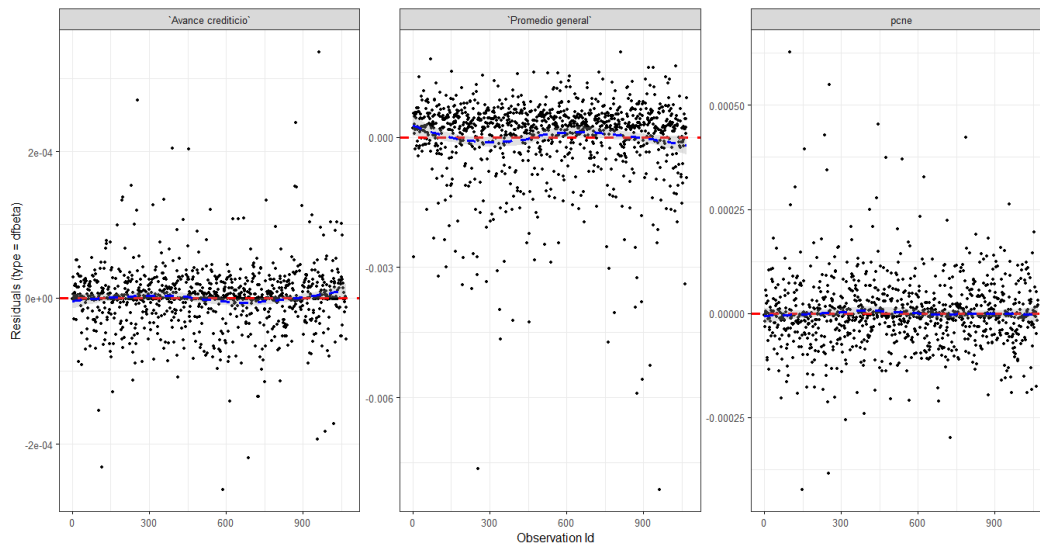


Figura 10. Prueba de observaciones influyentes

La figura 10 corresponde a los gráficos de residuos, en donde se evaluó la influencia de cada observación en los coeficientes del modelo de Cox, es decir, que si alguna observación fuera mayor a 0, tendría un fuerte impacto en los coeficientes del modelo. En este caso, todas las observaciones son próximas a 0, lo que indica que no cambian fuertemente a los coeficientes de la regresión.

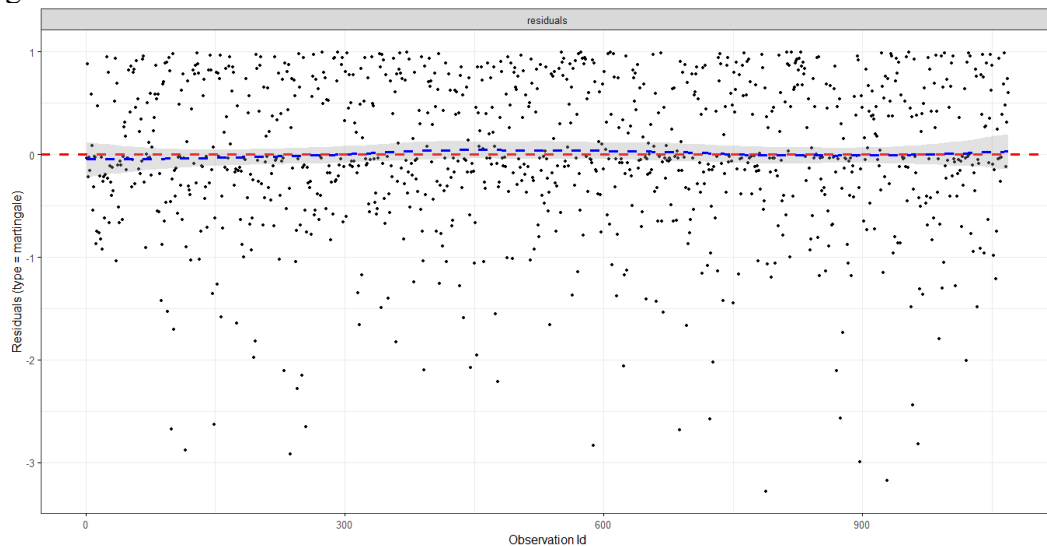


Figura 11. Residuos de Martingala

Mediante la figura 11 se corrobora el supuesto de no linealidad. Este gráfico también sirve para identificar datos atípicos. De acuerdo con lo observado, no se presentan patrones ni valores extremos, lo que indica que el modelo tiene un buen ajuste global y se cumple el

supuesto de no linealidad ya que se presenta una dispersión aleatoria. Esto mismo se comprobó con el test de Schoenfeld e indica que las estimaciones de los parámetros del modelo son confiables.

4.2.3. Evaluación del modelo

Para obtener la capacidad predictiva del modelo, se obtuvieron dos métricas.

La tabla 9 muestra el resultado obtenido para el índice de concordancia (C-index) el cual fue de 0.78, es decir, el modelo acertó en 78% de los pares concordantes e identificó y designó mayor riesgo al individuo que sufriría el evento de interés primero. Por otro lado, el 22% de los pares discordantes son el caso contrario a los pares concordantes.

Tabla 9. C-index e intervalos de confianza

C-index	0.78	
Intervalo de confianza	Inferior	Superior
Bootstrap al 95%	0.7534	0.8185

Una vez obtenido el índice de concordancia, se obtuvieron los intervalos de confianza al 95% generados mediante Bootstrap con 1,000,000 de réplicas sin reemplazo. Dado que el intervalo osciló entre 0.75 y 0.81, se concluye que el modelo generado es confiable y estadísticamente significativo ya que el valor obtenido en C-index (0.78) se encuentra dentro de este.

Tabla 10. Resultados de la puntuación de Brier (comparación de modelos)

Puntuación de Brier IBS [0; tiempo=1430)	
Modelo de referencia	0.246
Modelo de Cox	0.141
Modelo nulo	0.249

De acuerdo con los datos de la tabla 10, se obtuvo un IBS en el modelo de Cox de 0.141 que en comparación a los otros dos modelos generados (modelo de referencia y modelo nulo) el modelo de Cox propuesto tiene buen desempeño predictivo. Lo que confirma una vez más que el modelo propuesto es confiable y estadísticamente significativo.

5. Discusión y conclusiones

De acuerdo con los resultados obtenidos en el análisis exploratorio, la licenciatura con mayor porcentaje de egresados es Redes y Servicios de Cómputo. Este primer acercamiento con los datos también permitió identificar que la licenciatura en Ingeniería de Software es la que tiene mayor porcentaje de estudiantes que deciden desertar el nivel educativo.

El mejor método de estimación para el cálculo de los parámetros del modelo de regresión de Cox es el método de Verosimilitud Parcial Exacta ya que en comparación a los otros métodos de estimación, este obtuvo un menor AIC. Es indispensable observar el puntaje global del examen de admisión (EXANI-II, diseñado por CENEVAL, A.C.), el promedio general en licenciatura y el avance crediticio, siendo estas dos últimas las más significativas en el modelo. Mediante la verificación de los supuestos en regresión de Cox, especialmente en el test de residuos de Schoenfeld se comprobó que estas variables no violan el supuesto de proporcionalidad de riesgos ya que, al observar los valores de p de cada una de las variables y del modelo global, se concluye que no hay evidencia suficiente para rechazar la hipótesis nula, es decir que a lo largo del tiempo el riesgo se mantendrá constante.

El modelo es capaz de distinguir entre los individuos con menor y mayor riesgo de sufrir el evento de interés, esto se comprobó mediante la puntuación de Brier y el índice de concordancia. Lo que indica que con este modelo se pueden obtener resultados confiables y significativos.

Finalmente, a diferencia de lo reportado por Ramírez et al. (2017) se encontró que, para este caso, existen diferencias significativas entre los métodos de estimación ya que al comparar su AIC se obtuvieron valores diferentes, siendo el método de Breslow el que tuvo mayor AIC y Verosimilitud Parcial Exacta con un menor AIC.

En contraposición a lo reportado por Izasa et al. (2017), se encontró que las variables que influyen en la supervivencia académica son promedio general en licenciatura, avance crediticio y puntaje global obtenido en el examen de admisión, a diferencia del ingreso familiar y género reportado en dicho trabajo. De igual forma, Álvarez et al. (2024) reportaron que la sede, lugar y escuela de procedencia fueron las variables que explican la supervivencia académica, lo cual no coincide con los hallazgos de este trabajo.

Referencias

- Álvarez Carranza, A., Navarro Castillo, L. D., & González Jiménez, C. (2024). Análisis de supervivencia aplicado al tiempo de permanencia de los estudiantes de la facultad de ciencias económicas de la universidad hispanoamericana, 2021-2023. *Revista Boaciencia. Educación Y Ciencias Sociales*, 4(1), 1–24. <https://doi.org/10.59801/ecs.v4i1.239>
- Allende, A. S. (2024, 12 Marzo). Concordance Index as an Evaluation Metric - Analytics Vidhya - Medium. Medium. <https://medium.com/analytics-vidhya/concordance-index-72298c11eac7>
- Arguedas Negrini, I., & Jiménez Segura, F. (2007). Factores que promueven la permanencia de estudiantes en la educación secundaria. *Revista Electrónica "Actualidades Investigativas en Educación"*, 7(3), 0
- Arias, E. (2007). Étudiants et : un seul profil de réussite? <https://journals.openedition.org/pyramides/250?lang=en>
- Centurión-Cardena, H. J., Cano-Barrón, D. D., Sandoval-Gio, J., & Zapata-González, A. (2020). Redes neuronales para predecir el abandono académico en ingeniería. *Revista de Tecnología y Educación*, 1-7. <https://doi.org/10.35429/jtae.2020.11.4.1.7>
- Daza Vergaray, A. (2016). Un modelo basado en árboles de decisión para predecir la deserción estudiantil en la Educación Superior Privada. *UCV-Scientia*, 8(1), 59–73. <https://doi.org/10.18050/RevUcv-Scientia.v8n1a7>
- Keymolen, D. L., Arce, B. J. M., Linares, S. C. G. R., & Sánchez, Y. G. R. (2017). Dependencia funcional y su relación con la mortalidad general en adultos mayores. México: 2001-2015. *Población y Salud En Mesoamérica*, 15(2). <https://doi.org/10.15517/psm.v15i2.28660>
- Kuz, A., & Morales, R. (2023). Ciencia de Datos Educativos y aprendizaje automático: un caso de estudio sobre la deserción estudiantil universitaria en México. *Education in the Knowledge Society (EKS)*, 24, e30080. <https://doi.org/10.14201/eks.30080>
- Lazo, H. J. U., & Sánchez, Á. A. V. (2024). Aplicación del análisis de supervivencia para determinar el tiempo hasta la deserción escolar en la universidad. *Revista Cubana de Ciencias Informáticas*, 18(4), 9.
- M. Vergel-Ortega, J. J. Martínez-Lozano, E. Ibargüen-Mondragón, “Modelos estimados de análisis de supervivencia para el tiempo de permanencia de los estudiantes de la Universidad Francisco de Paula Santander”, *Respuestas*, vol. 21, no. 2, pp. 24-36, 2016.
- Martínez-González, Adrián, García-Minjares, Manuel, Buzo-Casanova, Enrique R., Gutiérrez-Cirlos, Carlos, Vargas-Castillo, José A., & Sánchez-Mendiola, Melchor. (2024). Trayectorias académicas como herramienta para identificar factores que afectan la regularidad en estudiantes de medicina. *Gaceta médica de México*, 160(6), 662-670. Epub 07 de febrero de 2025. <https://doi.org/10.24875/gmm.24000335>
- Montoya, J. R., Regino, E., & Gómez, S. Y. G. (2017). Comparación de métodos de estimación en regresión de Cox. *Comunicaciones en Estadística*, 10(1), 101-112.

- Murillo-Zabala, A. M., & De Los Santos, P. J. (2021). Permanencia estudiantil: factores que inciden en el Politécnico Internacional de Bogotá, Colombia. *Revista Electrónica Educare*, 25(1), 1-25. <https://doi.org/10.15359/ree.25-1.6>
- Narváez, Y. V., & Medina, M. A. G. (2017). Factores asociados a la permanencia de estudiantes universitarios: caso UAMM-UAT. *Revista de la Educación Superior*, 46(184), 117-138. <https://doi.org/10.1016/j.resu.2017.11.003>
- Olarte Moyano, J. C. (2020). Factores que predicen la permanencia estudiantil: análisis en un escenario de educación técnica. *CULTURA EDUCACIÓN Y SOCIEDAD*, 11(1), 25-38. <https://doi.org/10.17981/cultedusoc.11.1.2020.02>
- Osorio González, J. C. (2024). Comparación de modelos estadísticos y de aprendizaje automático para la predicción y prevención de la deserción estudiantil en los programas de pregrado de la Universidad EAFIT.
- Pérez, J., Chumacero, A. P., & Rodríguez, F. V. (2006). Los estudios de trayectoria escolar. Su aplicación en la educación media superior. *Graffylia*, 3(6), 24-29.
- Portal Martínez, Esther, Arias Fernández, Enrique, Lirio Castro, Juan, & Gómez Ramos, José Luís. (2022). Fracaso y abandono universitario: percepción de los(as) estudiantes de Educación social de la Universidad de Castilla La Mancha. *Revista mexicana de investigación educativa*, 27(92), 289-316. Epub 14 de marzo de 2022. Recuperado en 08 de abril de 2024, de http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-66662022000100289&lng=es&tlng=es.
- Puma Laura, C. V. (2018). Análisis de supervivencia de la deserción en los estudiantes de la facultad de Ciencias Naturales y Formales a través del Modelo de COX.
- Rochin Berumen, Fabiola Lydie. (2021). Deserción escolar en la educación superior en México: revisión de literatura. *RIDE. Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 11(22), e01. Epub 21 de mayo de 2021. <https://doi.org/10.23913/ride.v11i22.821>
- Salcedo, A. (2020). Deserción universitaria en Colombia. *Academia Y Virtualidad*, 3(1), 50-60. Recuperado a partir de <https://revistas.unimilitar.edu.co/index.php/ravi/article/view/5461>
- Saldarriaga-Isaza, C. A.; Tesorio, M.; Ochoa, G. I. (2019). Permanencia y deserción de estudiantes del programa PEAMA de la Universidad Nacional de Colombia - Sede Amazonia. *Mundo Amazónico*, 10(1): 245-259. <http://dx.doi.org/10.15446/ma.v10n1.66610>
- TRAYECTORIAS ESCOLARES EN LA UNIVERSIDAD VERACRUZANA. (s. f.-b). https://www.uv.mx/cpue/colped/N_2526/publragu.htm
- Vásquez Velásquez, J., Castaño Vélez, E., Gallón Gómez, S. A. y Gómez Portilla, K (2003). Determinantes de la deserción estudiantil en la Universidad de Antioquia. *Borradores del CIE*, (04), 1-38. Recuperado de <http://hdl.handle.net/10495/3627>
- Vergara Morales, J. R., Boj del Val, E., Barriga, O. A., & Díaz Larenas, C. (2017). Factores explicativos de la deserción de estudiantes de pedagogía. *Revista Complutense de Educación*, 2017, vol. 28, núm.. 2, p. 609-630

“Lis de Veracruz: Arte, Ciencia, Luz”

www.uv.mx

