

**UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO  
FACULTAD DE CIENCIAS QUÍMICAS, FÍSICAS Y MATEMÁTICAS  
ESCUELA PROFESIONAL DE MATEMÁTICA CON MENCIÓN EN  
ESTADÍSTICA**



**TESIS**

---

**DESISTIMIENTO ESTUDIANTIL EN LA ESCUELA PROFESIONAL DE  
MATEMÁTICA DE LA FACULTAD DE CIENCIAS QUÍMICAS, FÍSICAS Y  
MATEMÁTICAS-UNSAAC- PERIODO 2015-2023**

---

**PRESENTADO POR:**

Br. MARIO ANCALLA SIVINCHA

**PARA OPTAR AL TÍTULO  
PROFESIONAL DE LICENCIADO EN  
MATEMÁTICA MENCIÓN  
ESTADÍSTICA**

**ASESOR:**

Dr. EPIFANIO PUMA HUAÑEC

**CO-ASESOR:**

Mtro. CHARLES JHON MAMANI  
MAYTA

**CUSCO – PERÚ**

**2024**

# INFORME DE ORIGINALIDAD

(Aprobado por Resolución Nro. CU-303-2020-UNSAAC)

El que suscribe, asesor del trabajo de investigación/tesis titulado: Desistimiento estudiantil en la Escuela Profesional de Matemática de la Facultad de Ciencias Químicas, Físicas y Matemáticas - UNSAAC - periodo 2015-2023

presentado por: MARIO ANCALLA SIVINCHA

con Nro. de DNI: 44417100, para optar el título profesional/grado académico de LICENCIADO EN MATEMÁTICA MENCIÓN ESTADÍSTICA

Informo que el trabajo de investigación ha sido sometido a revisión por 1 veces, mediante el Software Antiplagio, conforme al Art. 6° del Reglamento para Uso de Sistema Antiplagio de la UNSAAC y de la evaluación de originalidad se tiene un porcentaje de 6%

Evaluación y acciones del reporte de coincidencia para trabajos de investigación conducentes a grado académico o título profesional, tesis

Porcentaje	Evaluación y Acciones	Marque con una (X)
Del 1 al 10%	No se considera plagio.	X
Del 11 al 30 %	Devolver al usuario para las correcciones.	
Mayor a 31%	El responsable de la revisión del documento emite un informe al inmediato jerárquico, quien a su vez eleva el informe a la autoridad académica para que tome las acciones correspondientes. Sin perjuicio de las sanciones administrativas que correspondan de acuerdo a Ley.	

Por tanto, en mi condición de asesor, firmo el presente informe en señal de conformidad y adjunto la primera hoja del reporte del Sistema Antiplagio.

Cusco, 23 de Octubre de 2024



Firma

Post firma Epifanio Roma Huáñec

Nro. de DNI 23861131

ORCID del Asesor 0000-0002-2592-7467

Se adjunta:

1. Reporte generado por el Sistema Antiplagio.
2. Enlace del Reporte Generado por el Sistema Antiplagio: oid: 27259:397682722

NOMBRE DEL TRABAJO

**TESIS - DESISTIMIENTO ESTUDIANTIL EN LA UNSAAC - MARIO ANCALLA SIVINCHA.pdf**

AUTOR

**MARIO ANCALLA SIVINCHA**

RECUENTO DE PALABRAS

**21328 Words**

RECUENTO DE CARACTERES

**121147 Characters**

RECUENTO DE PÁGINAS

**102 Pages**

TAMAÑO DEL ARCHIVO

**1.0MB**

FECHA DE ENTREGA

**Oct 23, 2024 2:30 PM GMT-5**

FECHA DEL INFORME

**Oct 23, 2024 2:32 PM GMT-5****● 6% de similitud general**

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cada base de datos.

- 5% Base de datos de Internet
- Base de datos de Crossref
- 4% Base de datos de trabajos entregados
- 1% Base de datos de publicaciones
- Base de datos de contenido publicado de Crossref

**● Excluir del Reporte de Similitud**

- Material bibliográfico
- Material citado
- Material citado
- Coincidencia baja (menos de 15 palabras)



## **DEDICATORIA**

A mis padres, Leónidas y Eustaquia; por haberme brindado la oportunidad de realizarme profesional, por su constante preocupación y apoyo indescriptible.

Con el mayor cariño, dedico el presente trabajo de investigación a mi querida esposa Norca, por ser siempre la constante compañera de exigencia, y nunca dudó en comprender y apoyar mis propósitos personales. A mis queridos hijos: Nharia, Leonard y Danhary, por ser fuente de mi permanente esfuerzo de superación, porque representan mi motivación constante.

## **AGRADECIMIENTO**

Mi agradecimiento a nuestro divino del alto, por permitir concretar este caro anhelo.

Mi absoluta muestra de gratitud al tan distinguido Dr. Epifanio Puma Huañec, por los grandes aportes y guías que han permitido concluir en la titulación.

Finalmente, mi profunda gratitud a mis maestros de la Escuela Profesional de Matemáticas de la facultad de Ciencias Físicas, Químicas y Matemáticas por haberme impartido sus excelsos conocimientos, y a la tricentenaria Universidad Nacional de San Antonio Abad de Cusco, por haberme cobijado en sus claustros del saber.

## ÍNDICE

DEDICATORIA .....	ii
AGRADECIMIENTO .....	iii
ÍNDICE DE TABLAS .....	viii
ÍNDICE DE FIGURAS .....	x
PRESENTACIÓN .....	xi
RESUMEN .....	xii
ABSTRACT .....	xiii
INTRODUCCIÓN.....	14
CAPÍTULO I PLANTEAMIENTO DEL PROBLEMA.....	16
1.1.    Situación problemática .....	16
1.2.    Formulación del problema.....	19
1.2.1.    Problema general .....	19
1.2.2.    Problemas específicos.....	19
1.3.    Justificación de la investigación.....	19
1.4.    Objetivos de la investigación.....	21
1.4.1.    Objetivo general .....	21
1.4.2.    Objetivos específicos.....	21
CAPÍTULO II MARCO TEÓRICO.....	23
2.1.    Bases teóricas .....	23
2.1.1.    Análisis de supervivencia .....	23
2.1.2.    Definición de términos empleados en el análisis de supervivencia.....	24

2.1.3.	Censura y sus tipos .....	26
2.1.4.	Probabilidades de supervivencia .....	29
2.1.5.	Funciones básicas de supervivencia .....	31
2.1.6.	Estimador Kaplan Meier.....	35
2.1.7.	Estimador de Nelson Alaen .....	36
2.1.8.	Prueba de hipótesis para la función de supervivencia .....	36
2.1.9.	Modelo de Cox .....	40
2.2.	Marco conceptual .....	52
2.1.1.	Desistimiento .....	52
2.2.2.	Desistimiento Universitario.....	53
2.2.3.	Deserción estudiantil .....	54
2.2.4.	Factores que contribuyen a la deserción estudiantil .....	55
2.2.5.	Impacto de la deserción estudiantil .....	56
2.2.6.	Medidas para prevenir la deserción estudiantil .....	57
2.3.	Antecedentes de estudio .....	58
2.3.1.	Antecedentes internacionales .....	58
2.3.2.	Antecedentes nacionales.....	60
2.3.3.	Antecedentes locales.....	61
CAPÍTULO III HIPOTESIS Y VARIABLES .....		63
3.1.	Hipótesis .....	63
3.1.1.	Hipótesis general .....	63
3.1.2.	Hipótesis específicas.....	63

3.2.	Identificación de variables e indicadores.....	63
3.3.	Operacionalización de variables.....	65
CAPÍTULO IV METODOLOGÍA.....		67
4.1.	Ámbito de estudio: Localización política y geográfica.....	67
4.2.	Tipo y nivel de la investigación.....	67
4.3.	Método de investigación.....	68
4.4.	Diseño de investigación.....	68
4.5.	Unidad de Análisis .....	68
4.6.	Población y muestra .....	68
4.7.	Técnicas e instrumentos de recolección de datos .....	69
4.8.	Técnicas de análisis e interpretación de la información .....	70
CAPÍTULO V RESULTADOS Y DISCUSIÓN .....		71
5.1.	Procesamiento, análisis e interpretación.....	71
5.1.1	Análisis exploratorio de datos .....	71
5.1.2.	Resultado de la estimación de la función de supervivencia mediante el método de Kaplan-Meier .....	77
5.1.3.	Resultado de la estimación de la función de riesgo acumulado mediante el método de Nelson-Aalen .....	80
5.1.4.	Resultados de la comparación de las funciones de supervivencia según factores.....	82
5.1.5.	Resultado del modelo de regresión de Cox .....	87
5.2.	Discusiones.....	94

CONCLUSIONES.....	96
RECOMENDACIONES .....	97
BIBLIOGRAFÍA .....	99
ANEXOS .....	106
A. Matriz de consistencia .....	107
B. Tiempo de desistimiento y los semestres académicos (2015-2023).....	108
C. Data utilizada para la investigación.....	109

## ÍNDICE DE TABLAS

<b>Tabla 1</b> Prueba de igualdad de la función de supervivencia de M grupos de observación .....	39
<b>Tabla 2</b> Matriz de operacionalización de variables .....	65
<b>Tabla 3</b> Sexo de los estudiantes matriculados durante 2015-2023 EP. Matemática UNSAAC .....	71
<b>Tabla 4</b> Condición Económica de los estudiantes matriculados durante 2015-2023 EP. Matemática UNSAAC .....	72
<b>Tabla 5</b> Edad de los estudiantes matriculados durante el periodo de estudio de la E.P Matemática de la UNSAAC Cusco .....	75
<b>Tabla 6</b> Numero de hermanos de los estudiantes matriculados durante el periodo de estudio de la E.P Matemática de la UNSAAC .....	75
<b>Tabla 7</b> Resumen para la variable evento de los estudiantes matriculados durante el periodo de estudio de la E.P Matemática de la UNSAAC .....	76
<b>Tabla 8</b> Los tiempos de supervivencia según el método de Kaplan-Meier por semestre académico de los estudiantes matriculados durante el periodo 2015-2023 de la E.P Matemática de la UNSAAC .....	77
<b>Tabla 9</b> Resultados de la prueba de Log-Rank para el sexo de los estudiantes matriculados en la EPM durante 2015-2023 .....	83
<b>Tabla 10</b> Comparaciones globales .....	84
<b>Tabla 11</b> Resultados de la prueba de Log-Rank para la condición económica de los estudiantes matriculados durante el periodo 2015-2023 de la E.P Matemática de la UNSAAC .....	86
<b>Tabla 12</b> Comparaciones globales .....	86

<b>Tabla 13</b> Modelos de regresión de Cox .....	88
<b>Tabla 14</b> Coeficientes de regresión de Cox para las variables del modelo .....	90
<b>Tabla 15</b> Comparaciones globales .....	92
<b>Tabla 16</b> Contraste de hipótesis del modelo de Cox .....	92

## ÍNDICE DE FIGURAS

<b>Figura 1</b> Sexo de los estudiantes matriculados durante 2015-2023 E.P. Matemática UNSAAC.....	71
<b>Figura 2</b> Condición Económica de los estudiantes matriculados durante 2015-2023 EP. Matemática UNSAAC.....	73
<b>Figura 3</b> Lugar de procedencia de los estudiantes matriculados durante el periodo de estudio de la E.P Matemática de la UNSAAC .....	74
<b>Figura 4</b> Curva de supervivencia de Kaplan-Meier- los estudiantes matriculados durante el periodo 2015-2023 de la E.P Matemática de la UNSAAC .....	79
<b>Figura 5</b> Curva del riesgo acumulado según Nelson-Aalen - estudiantes matriculados durante el periodo 2015-2023 de la E.P Matemática de la UNSAAC .....	80
<b>Figura 6</b> Curva de supervivencia de los estudiantes matriculados durante el periodo 2015-2023 de la E.P Matemática de la UNSAAC según sexo.....	82
<b>Figura 7</b> Curva de supervivencia de los estudiantes matriculados durante el periodo 2015-2023 de la E.P Matemática de la UNSAAC según condición económica .....	85

## **PRESENTACIÓN**

SEÑOR DECANO DE LA FACULTAD DE CIENCIAS QUÍMICAS, FÍSICAS Y MATEMÁTICAS,

SEÑOR DIRECTOR DE LA ESCUELA PROFESIONAL DE MATEMÁTICA,

SEÑORES DOCENTES MIEMBROS DEL JURADO:

En cumplimiento con las normas de grados y títulos de la Escuela Profesional de Matemática con mención en Estadística, les presentamos para su evaluación el trabajo titulado DESISTIMIENTO ESTUDIANTIL EN LA ESCUELA PROFESIONAL DE MATEMÁTICA DE LA FACULTAD DE CIENCIAS QUÍMICAS, FÍSICAS Y MATEMÁTICA-UNSAAC- PERIODO 2015-2023 con el fin de optar al Título Profesional de Licenciado en Matemática Mención Estadística.

Esta investigación aborda el desistimiento estudiantil. Para su análisis, se aplicaron diversas técnicas de análisis de supervivencia, que permiten estudiar el tiempo hasta que ocurre un evento de interés, en este caso, el desistimiento de los estudiantes. En particular, se ha utilizado el modelo de riesgos proporcionales de Cox, que permite identificar los factores que incrementan o disminuyen la probabilidad de que un estudiante abandone la carrera. Adicionalmente, se empleó el estimador Kaplan-Meier, el estimador de Nelson-Aalen y la prueba de Log-Rank.

El objetivo final de esta investigación es proporcionar información que sirva para la toma de decisiones orientadas a mejorar las tasas de retención y apoyar a los estudiantes en riesgo de abandonar la carrera, con el fin de contribuir al fortalecimiento del sistema educativo en nuestra universidad.

Atentamente,

Br. Mario Ancalla Sivincha

## RESUMEN

La presente investigación se centra en el análisis de los factores asociados al desistimiento estudiantil en la Escuela Profesional de Matemática de la Universidad Nacional de San Antonio Abad del Cusco (UNSAAC) durante el periodo 2015-2023. Los objetivos incluyen identificar los factores que inciden en la probabilidad de abandono académico, estimar la función de supervivencia y aplicar un modelo de riesgos proporcionales de Cox para determinar los elementos determinantes en el desistimiento. La metodología adoptada abarca desde un análisis exploratorio de datos hasta la implementación de técnicas estadísticas avanzadas, como el estimador de Kaplan-Meier para la función de supervivencia y el modelo de riesgos proporcionales de Cox para identificar los factores que afectan el abandono. El análisis de los datos se realizó utilizando el software R. Los resultados indican que existen múltiples factores que influyen en el desistimiento estudiantil. La estimación de la función de supervivencia sugiere que los estudiantes en riesgo de abandonar su formación presentan características socioeconómicas y académicas específicas. Además, el modelo de riesgos proporcionales de Cox valida que factores como la edad, el rendimiento académico y la condición socioeconómica son determinantes significativos en la probabilidad de abandono. La comparación de las curvas de supervivencia revela que los estudiantes de edades más jóvenes tienen mayores probabilidades de permanecer en la carrera. Estos hallazgos destacan la necesidad de implementar políticas de apoyo académico y emocional para los estudiantes en riesgo, así como fortalecer los programas de tutoría para mejorar la retención en la carrera de Matemática.

**Palabras Clave:** *desistimiento estudiantil, análisis de supervivencia, factores socioeconómicos, rendimiento académico.*

## ABSTRACT

This research focuses on the analysis of the factors associated with student dropout in the Professional School of Mathematics at the National University of San Antonio Abad del Cusco (UNSAAC) during the period 2020-2023. The objectives include identifying the factors that affect the probability of academic dropout, estimating the survival function and applying a Cox proportional hazards model to determine the determinants of dropout. The methodology adopted ranges from exploratory data analysis to the implementation of advanced statistical techniques, such as the Kaplan-Meier estimator for the survival function and the Cox proportional hazards model to identify the factors affecting dropout. Data analysis was performed using R software. The results indicate that there are multiple factors influencing student dropout. The estimation of the survival function suggests that students at risk of dropping out have specific socio-economic and academic characteristics. In addition, the Cox proportional hazards model validates that factors such as age, academic performance and socio-economic status are significant determinants of the probability of dropping out. Comparison of the survival curves reveals that students of younger ages are more likely to stay in school. These findings highlight the need to implement academic and emotional support policies for at-risk students, as well as to strengthen tutoring programmes to improve retention in Mathematics.

Keywords: student dropout, survival analysis, socioeconomic factors, academic performance.

## INTRODUCCIÓN

A nivel internacional, el desistimiento estudiantil es una problemática de gran relevancia en las instituciones de educación superior. Según varios estudios (Allen y Seaman, 2015; Tinto, 2017), países como Estados Unidos, Canadá y varias naciones europeas han reportado altos índices de desistimiento en carreras de ciencias, tecnología, ingeniería y matemáticas (STEM, por sus siglas en inglés). Diversos investigadores han señalado que factores como la falta de adaptación a la vida universitaria, el nivel académico de los programas y las dificultades financieras son los principales detonantes del abandono (Bean, 1980; Kuh et al., 2006). En este contexto, el desistimiento no solo afecta a los estudiantes, sino también a las universidades, al reducir su número de graduados, y a las economías, que enfrentan un déficit de profesionales cualificados en áreas estratégicas para el desarrollo (Tinto, 1993).

En América Latina, el desistimiento universitario es un tema de constante preocupación, ya que se asocia no solo a factores académicos, sino también a la desigualdad socioeconómica que caracteriza a la región. Según el Banco Interamericano de Desarrollo (BID, 2020), en países como Perú las tasas de abandono universitario han aumentado en los últimos años, especialmente en carreras científicas. Informes nacionales del Ministerio de Educación (MINEDU, 2021) revelan que, si bien el acceso a la educación superior ha mejorado, la permanencia de los estudiantes sigue siendo un desafío. Entre los principales factores que afectan el éxito académico en Perú se encuentran la falta de orientación vocacional adecuada, las dificultades económicas y la baja calidad de la educación previa a la universidad, como lo señalan investigaciones de De la Torre y Vargas (2019).

En el ámbito local, específicamente en la ciudad de Cusco, la situación no es diferente. La Escuela Profesional de Matemática, como parte de la Universidad Nacional

de San Antonio Abad del Cusco (UNSAAC), enfrenta retos similares en cuanto a la permanencia estudiantil. Un informe interno de la universidad (UNSAAC, 2023) muestra que, pese a la calidad de la enseñanza, un porcentaje significativo de estudiantes no logra culminar sus estudios, lo que refleja un problema multifactorial. Las particularidades geográficas, socioeconómicas y culturales de la región influyen en la decisión de muchos estudiantes de abandonar sus estudios. Según datos de la Dirección Regional de Educación Cusco (DREC, 2022), la lejanía de los distritos rurales, los costos asociados a la vida universitaria y la presión académica de una carrera como Matemática son factores que impactan negativamente en la retención estudiantil.

La relevancia de este estudio radica en que busca proporcionar un análisis profundo de los factores que contribuyen al desistimiento estudiantil en la Escuela Profesional de Matemática de la UNSAAC, utilizando técnicas de análisis de supervivencia. Este enfoque permitirá entender no solo qué porcentaje de estudiantes abandona sus estudios, sino también cuándo lo hacen y bajo qué circunstancias (Tinto, 2006). El objetivo es ofrecer recomendaciones basadas en evidencia para mejorar las políticas de retención estudiantil en esta carrera, lo que podría tener un impacto significativo en el éxito académico y profesional de los futuros matemáticos de la región.

La estructura de esta tesis se organiza de la siguiente manera: en el primer capítulo, se presentará el marco teórico, donde se revisarán estudios previos relacionados con el desistimiento estudiantil y las metodologías empleadas para su análisis. En el segundo capítulo, se describirá la metodología utilizada en este estudio, con énfasis en el modelo de análisis de supervivencia aplicado. El tercer capítulo presentará los resultados obtenidos y su interpretación, mientras que el cuarto y último capítulo estará dedicado a las conclusiones y recomendaciones, basadas en los hallazgos del análisis (adaptado de Creswell, 2014).

## CAPÍTULO I

### PLANTEAMIENTO DEL PROBLEMA

#### 1.1. Situación problemática

En la Escuela Profesional de Matemática de la Facultad de Ciencias Químicas, Físicas y Matemáticas en la UNSAAC, es un tema de discusión permanente el desistimiento, deserción o evasión estudiantil universitaria en la escuela profesional de Matemática, este desistimiento se entiende como la disolución del vínculo que tiene el estudiante a través del proceso la matrícula académica en un semestre determinado o abandono de los estudios universitarios, este hecho trae como consecuencia efectos de tipo financiero, académico y social que implican la pérdida de esfuerzos y recursos en el país.

Según el Instituto Nacional de Estadística e Informática (INEI), la tasa de deserción universitaria en Perú es del 14,5%. Esto significa que, de cada 100 estudiantes que ingresan a la universidad, 14,5 no terminan sus estudios, sin embargo, no se tiene información sobre la deserción de estudiantes universitarios que estudian matemática.

El presupuesto asignado a la educación universitaria en el año 2023, según el Ministerio de Educación del Perú, fue de S/ 2.137 millones, este presupuesto representa el 4,6% del presupuesto total del sector educación de esto, se destina Remuneraciones y pensiones: S/ 1.203 millones, Intervenciones y educación superior: S/ 734 millones, Infraestructura: S/ 190 millones y Gestión: S/ 10 millones. El costo estimado que se destina a cada estudiante universitario en el año 2023 en Perú es de S/ 577.158. Este costo se calcula dividiendo el presupuesto total en educación universitaria por la matrícula total de estudiantes universitarios, en el año 2023, la matrícula total de estudiantes universitarios en Perú fue de 2.276.789. Por lo tanto, el costo estimado por estudiante universitario es de:  $S/ 2.137 \text{ millones} / 2.276.789 \text{ estudiantes} = S/ 577.158$

Por lo que el desistimiento estudiantil universitaria tiene un efecto negativo en la: Universidad: El desistimiento afecta la reputación y la sostenibilidad financiera de las universidades. Las universidades que tienen altas tasas de desistimiento pueden ser percibidas como instituciones de baja calidad. Además, las universidades pueden perder ingresos por matrícula si los estudiantes abandonan sus estudios antes de completarlos.

Impacto en la sociedad: El desistimiento también puede contribuir al aumento de la desigualdad social. Los estudiantes que abandonan la universidad tienen menos probabilidades de encontrar un trabajo bien remunerado y de tener éxito en su carrera. Esto puede contribuir a la brecha entre los ricos y los pobres.

Impacto en los estudiantes: Para los estudiantes, el desistimiento puede significar la pérdida de oportunidades de desarrollo personal y profesional. Los estudiantes que abandonan la universidad tienen menos probabilidades de encontrar un trabajo bien remunerado y de tener éxito en su carrera. Además, el desistimiento puede tener un impacto negativo en la autoestima y el bienestar de los estudiantes.

El informe de la Evaluación Censal de Estudiantes ECE 2022 muestra que, en Matemática, el 49% de los estudiantes de quinto grado de secundaria no alcanzaron el nivel esperado de desempeño. Los resultados también revelan que existen brechas de rendimiento significativas entre regiones y zonas rurales y urbanas.

Impacto en las familias: El desistimiento también puede representar un gasto económico significativo para las familias. Los estudiantes universitarios suelen tener que asumir gastos personales, como alimentación, transporte, alojamiento, y otros gastos imprevistos. Si un estudiante abandona la universidad, sus padres pueden tener que asumir estos gastos por un período de tiempo más largo.

El desistimiento estudiantil es un problema que afecta a las universidades de todo el mundo. En el Perú, el desistimiento estudiantil en el nivel universitario es de alrededor del 30%. En la Universidad Nacional de San Antonio Abad del Cusco (UNSAAC), el desistimiento estudiantil en la escuela profesional de Matemática es de alrededor del 60%.

Este problema tiene un impacto negativo en los estudiantes, en la universidad y en la sociedad en general. Los estudiantes que desertan tienen menos probabilidades de obtener un título universitario, lo que puede dificultar su acceso al mercado laboral y a oportunidades de desarrollo personal y profesional. La universidad pierde estudiantes potenciales, lo que puede afectar su reputación y su capacidad de cumplir con su misión. La sociedad en general pierde recursos humanos capacitados, lo que puede afectar el desarrollo económico y social.

Según la ENAHO 2022, la tasa de desistimiento estudiantil en el nivel universitario en el Perú es de 29,9%. Esta tasa es mayor en las universidades públicas, donde alcanza el 32,6%.

En la UNSAAC, la tasa de desistimiento estudiantil en la escuela profesional de Matemática es de 61,8%. Esta tasa es mayor en los primeros años de la carrera, donde alcanza el 75,5%, así mismo el promedio de estudiantes que ingresan a la escuela profesional de matemática es de 190 estudiantes por semestre.

Los resultados del presente trabajo de investigación permitirán identificar los factores que contribuyen al desistimiento estudiantil en la escuela profesional de Matemática de la UNSAAC. Esta información puede ser utilizada por la universidad para desarrollar medidas para prevenir el desistimiento estudiantil.

## **1.2. Formulación del problema**

### **1.2.1. Problema general**

¿Cómo se puede explicar el desistimiento estudiantil en la Escuela Profesional de Matemática de la UNSAAC durante el periodo 2015-2023 utilizando el modelo de riesgos proporcionales de Cox?

### **1.2.2. Problemas específicos**

- a) ¿Cuál es la probabilidad de que los estudiantes de la Escuela Profesional de Matemática continúen sus estudios a lo largo del tiempo durante el periodo 2015-2023?
- b) ¿Cómo se acumula el riesgo de desistimiento a medida que transcurre el tiempo durante el periodo 2015-2023?
- c) ¿Existen diferencias significativas en las curvas de supervivencia del desistimiento estudiantil según variables categóricas como el sexo, el lugar de procedencia o la condición socioeconómica?

## **1.3. Justificación de la investigación**

El desistimiento estudiantil es un problema grave que afecta a las universidades de todo el mundo. En el Perú, la tasa de desistimiento estudiantil en la educación superior es del 30%, lo que significa que tres de cada diez estudiantes abandonan la universidad antes de completar su carrera.

El desistimiento estudiantil tiene un impacto negativo en los estudiantes, las universidades y la sociedad en general. Los estudiantes que desertan tienen menos probabilidades de encontrar un empleo bien remunerado, lo que puede conducir a la pobreza y la exclusión social. Las universidades pierden recursos y prestigio cuando los

estudiantes desertan, y la sociedad pierde la oportunidad de contar con profesionales bien capacitados.

El desistimiento estudiantil puede estar causada por una variedad de factores, incluyendo factores económicos, académicos, personales e institucionales. Los factores económicos incluyen la falta de recursos económicos para pagar la matrícula, los libros y otros gastos relacionados con la educación. Los factores académicos incluyen las dificultades académicas, el fracaso escolar y la falta de motivación para estudiar. Los factores personales incluyen problemas familiares, de salud o de pareja. Los factores institucionales incluyen la calidad de la educación, la falta de apoyo a los estudiantes y la burocracia.

Los factores que contribuyen al desistimiento estudiantil en la escuela profesional de Matemática de la Facultad de Ciencias Químicas, Físicas y Matemática-UNSAAC-periodo 2020-2023, se basa en los siguientes datos y fuentes:

Datos estadísticos: Según datos de la Dirección General de Educación Superior Universitaria (DIGESU) del Ministerio de Educación del Perú, la tasa de desistimiento estudiantil en la carrera de Matemática de la UNSAAC en el periodo 2020-2023 fue del 30%. Esto significa que, de cada 100 estudiantes que ingresaron a la carrera en 2020, 30 abandonaron la carrera antes de completarla.

Estudios previos realizado sobre los Factores que contribuyen al desistimiento estudiantil en la carrera de Matemática de la Universidad Nacional de Trujillo por la Universidad Nacional de Trujillo en 2022 publicado en la Revista de Investigación Educativa, Vol. 10, No. 1, pp. 1-15, 2022 por los Autores: José Luis Medina-Muñoz, Luis Enrique López-Reynoso, y José Antonio Paredes-Reyes encontró que los factores que

contribuyen al desistimiento estudiantil en la carrera de Matemática de esa universidad son los siguientes:

- a) **Factores económicos:** La falta de recursos económicos para pagar la matrícula, los libros y otros gastos relacionados con la educación.
- b) **Factores académicos:** Las dificultades académicas, el fracaso escolar y la falta de motivación para estudiar.
- c) **Factores personales:** Problemas familiares, de salud o de pareja.
- d) **Factores institucionales:** La falta de apoyo a los estudiantes por parte de la universidad.

El estudio propuesto contribuirá a la determinación del porcentaje de estudiantes que desistieron en la escuela profesional de matemática por semestre en el periodo 2015-2023 e identificación de los factores que contribuyen al desistimiento estudiantil en la escuela profesional de Matemática de la UNSAAC. Los resultados del estudio podrán utilizarse para tomar medidas que permitan prevenir el desistimiento estudiantil en esta escuela profesional.

#### **1.4.Objetivos de la investigación**

##### ***1.4.1. Objetivo general***

Explicar el desistimiento estudiantil en la Escuela Profesional de Matemática de la UNSAAC durante el periodo 2015-2023 utilizando el modelo de riesgos proporcionales de Cox.

##### ***1.4.2. Objetivos específicos***

- a) Estimar la probabilidad de que los estudiantes continúen sus estudios a lo largo del tiempo durante el periodo 2015-2023.

- b) Estimar la acumulación del riesgo de desistimiento estudiantil conforme transcurre el tiempo durante el periodo 2015-2023.
- c) Comparar las curvas de supervivencia del desistimiento estudiantil según variables categóricas como el sexo, el lugar de procedencia y la condición socioeconómica.

## **CAPÍTULO II**

### **MARCO TEÓRICO**

#### **2.1. Bases teóricas**

##### **2.1.1. Análisis de supervivencia**

El análisis de supervivencia es un conjunto de técnicas estadísticas diseñadas para analizar datos en los que el resultado es el tiempo hasta que ocurre un evento de interés. Este tipo de análisis es ampliamente utilizado en la biomedicina, particularmente en estudios de seguimiento para determinar la eficacia de tratamientos y la supervivencia de los pacientes (Hosmer et al., 2008). La importancia de este análisis radica en su capacidad para manejar datos censurados, que son comunes en estudios donde los individuos no experimentan el evento dentro del tiempo de observación.

El principal objetivo del análisis de supervivencia es describir y modelar la distribución del tiempo hasta que se produzca un evento. Entre los modelos más utilizados se encuentran el modelo de riesgos proporcionales de Cox y el estimador de Kaplan-Meier, que permiten ajustar por múltiples covariables o comparar grupos, respectivamente (Hosmer et al., 2008). Este tipo de análisis no solo es útil en medicina, sino que también se aplica en áreas como la ingeniería, para el análisis de fiabilidad de sistemas, o en ciencias sociales, para estudiar eventos como la duración del matrimonio o el desempleo (Kleinbaum y Klein, 2012).

Los datos de supervivencia tienen características particulares que los distinguen de otros tipos de datos. La censura y el hecho de que el evento puede no ocurrir para todos los individuos son factores clave que hacen necesario el uso de métodos especializados. Este tipo de análisis puede ofrecer información valiosa

sobre cómo influyen distintas covariables en la ocurrencia del evento, permitiendo, por ejemplo, identificar factores de riesgo en enfermedades o predecir fallos en sistemas mecánicos (Alvarez, 2023).

### **2.1.2. Definición de términos empleados en el análisis de supervivencia**

- a) **Fecha inicial:** Este término se refiere al momento en que se comienza a observar a un individuo o sujeto dentro de un estudio. La fecha inicial es crucial, ya que marca el inicio del seguimiento y el tiempo desde el cual se empieza a contabilizar la supervivencia del individuo hasta que ocurre el evento de interés o hasta que se alcanza el final del estudio. Según Kleinbaum y Klein (2012), esta fecha puede variar entre individuos si se trata de estudios observacionales donde la entrada al estudio no es simultánea para todos los sujetos, y puede ser el momento de diagnóstico o de inclusión en el estudio.
- b) **Fecha de última noticia:** La fecha de última noticia se refiere al momento más reciente en el cual se tiene información sobre el estado de un individuo dentro del estudio. Esto es especialmente importante en estudios de larga duración o con seguimiento irregular, ya que puede haber periodos en los que no se recojan datos sobre ciertos participantes. Según Rubio y Martínez (2016), esta fecha es utilizada para calcular el tiempo de seguimiento efectivo de un individuo cuando el evento de interés no ha ocurrido. En estudios con censura, esta fecha ayuda a determinar si el individuo sigue en riesgo o si su seguimiento ha concluido sin observar el evento.
- c) **Seguimiento:** El seguimiento es el proceso mediante el cual se monitorea a los individuos a lo largo del tiempo para registrar si experimentan el evento de interés. El seguimiento puede ser continuo o en intervalos y es un aspecto esencial para determinar los tiempos de supervivencia. Según Salazar et al. (2020), el seguimiento efectivo es vital para garantizar que los datos reflejen adecuadamente la probabilidad

de ocurrencia del evento y la relación con posibles covariables. El seguimiento también afecta directamente la estimación de la función de supervivencia, ya que permite observar tanto los tiempos hasta el evento como los casos censurados.

- d) **Periodo de seguimiento:** Este término se refiere al lapso de tiempo durante el cual se observa a los individuos en un estudio. El periodo de seguimiento puede variar entre individuos, dependiendo de cuándo ingresaron al estudio y de la duración del estudio en sí. Según Klein et al. (2014) el periodo de seguimiento es un factor crucial en los análisis de supervivencia, ya que determina la cantidad de información disponible para cada individuo y afecta la precisión de las estimaciones. Un periodo de seguimiento insuficiente puede llevar a una alta proporción de censura, lo que afecta la validez de los resultados.
- e) **Fecha de finalización del estudio:** Este es el momento en el cual se concluye formalmente el estudio, y se detiene la recolección de datos. La fecha de finalización marca el fin del periodo de observación para todos los individuos, ya sea que hayan experimentado el evento de interés o no. Klein et al. (2014) explican que, en estudios con censura a la derecha, muchos individuos pueden no haber experimentado el evento para cuando se alcanza esta fecha, lo que introduce la necesidad de un análisis adecuado de los datos censurados.
- f) **Tiempo de supervivencia:** El tiempo de supervivencia es el intervalo transcurrido desde un punto inicial hasta la ocurrencia del evento de interés, como la muerte, la recaída o la falla de un sistema. Según Rossello y González (2022), este tiempo es el principal resultado de interés en los estudios de supervivencia y se modela utilizando funciones como la de supervivencia  $S(t)$  o la función de riesgo  $h(t)$ . El tiempo de supervivencia puede estar influenciado por múltiples factores, como el tratamiento

recibido, las características individuales o el entorno, lo que lo convierte en una variable compleja y multidimensional.

- g) **Observación censurada:** Se refiere a los casos en los que no se observa el tiempo exacto en que ocurre el evento de interés. La censura puede darse porque el evento no ocurre antes del final del seguimiento o porque el individuo sale del estudio antes de tiempo por razones ajenas al evento de interés. Según Vega (2019), las observaciones censuradas son un componente común en los estudios de supervivencia y requieren técnicas específicas para ser analizadas correctamente. Los datos censurados aportan información valiosa, ya que indican que el evento no ha ocurrido hasta un determinado punto en el tiempo.

### 2.1.3. Censura y sus tipos

El concepto de censura es fundamental en el análisis de supervivencia, donde no todos los eventos de interés son observados dentro del período de seguimiento. Esto resulta en datos incompletos sobre el tiempo hasta que ocurre un evento de interés (como la muerte o el fracaso de un componente), afectando a una proporción considerable de la muestra en estudio. Este fenómeno plantea desafíos significativos para la inferencia estadística en estudios de supervivencia y fiabilidad, ya que impide la disponibilidad de información completa sobre todos los individuos en la muestra (Kleinbaum y Klein, 2012)

En análisis de supervivencia, la censura ocurre cuando el tiempo completo de supervivencia no se conoce para todos los individuos observados. Las razones de la censura pueden variar, incluyendo el final del período de estudio antes de que el evento ocurra, o la pérdida de seguimiento de los participantes por razones diversas como la salida del estudio o la pérdida de contacto. Estos elementos introducen una complejidad

adicional en el análisis, ya que los métodos estadísticos convencionales asumen típicamente datos completos (Hosmer et al., 2008).

Para gestionar la censura de manera efectiva, es crucial entender sus tipos y cómo impactan en el análisis de los datos. Se identifican tres categorías principales de censura: censura por la derecha, censura por la izquierda y censura por intervalo. La más común es la censura por la derecha, donde el tiempo hasta el evento es mayor que el tiempo de seguimiento. La censura por la izquierda ocurre cuando el evento sucede antes de que el individuo entre al estudio, y la censura por intervalo se refiere a casos donde el evento ocurre dentro de un intervalo de tiempo específico, pero el momento exacto es desconocido (Martín, 2022).

#### **a) Censura por derecha**

En el contexto del análisis de supervivencia, la censura por la derecha ocurre cuando no se ha observado el evento de interés, como la muerte o la falla de un equipo, dentro del período de seguimiento del estudio. Este tipo de censura es identificado cuando el tiempo hasta el evento real es superior al tiempo de seguimiento registrado, lo que resulta en una falta de información completa sobre el tiempo exacto del evento (Martín, 2022).

La censura por la derecha es el tipo más frecuente de censura en estudios de supervivencia debido a que los períodos de seguimiento suelen ser fijos y no todos los individuos experimentan el evento antes de que el estudio concluya. Además, algunos participantes pueden salir del estudio por razones no relacionadas con el evento en estudio, como cambios en su situación personal o salud, complicando aún más la recopilación de datos completos sobre el tiempo hasta el evento (Hosmer et al., 2008).

Desde un punto de vista metodológico, manejar la censura por la derecha involucra ajustar las estimaciones para considerar que el tiempo hasta el evento para algunos individuos es desconocido, pero excede el tiempo observado. Los análisis estadísticos

deben incorporar métodos que adecuen estas incertidumbres, como el uso del estimador de Kaplan-Meier o modelos de riesgos proporcionales de Cox, que son diseñados para generar inferencias precisas bajo condiciones de datos censurados (Ortega et al., 2023)

De forma general, toda variable censurada por la derecha, se define como:

$$Y_i = \begin{cases} k_i, & \text{si } y_i > k_i \\ y_i, & \text{si } y_i \leq k_i \end{cases}$$

Note que  $P(Y_i = k_i) = P(Y_i > k_i) \geq 0$ , ya que  $Y \in \mathbb{R}$ . Luego,  $Y_i = \min\{y_i, k_i\}, \forall i$ .

### **b) Censura por izquierda**

La censura por la izquierda ocurre cuando el evento de interés ha sucedido antes del inicio del periodo de observación del estudio. En estos casos, la información sobre el momento exacto del evento es desconocida, aunque se sabe que ocurrió antes de un punto temporal específico. Esto es particularmente común en estudios retrospectivos donde se recopilan datos de eventos que sucedieron antes de que los sujetos fueran incluidos en el estudio (Kleinbaum y Klein, 2012).

Este fenómeno presenta desafíos únicos para el análisis de datos, ya que implica que el tiempo real hasta el evento es más corto que el tiempo observado, y este subregistro puede sesgar los análisis si no se maneja adecuadamente. Los métodos estadísticos convencionales, que presuponen datos completos y precisos sobre los tiempos de evento, pueden no ser adecuados en estas circunstancias (Hernández Z. , 2012).

Para abordar la censura por la izquierda, los investigadores deben emplear técnicas que permitan incorporar esta información limitada de manera efectiva en sus modelos estadísticos. Estrategias como la modelización de datos truncados o el uso de métodos de máxima verosimilitud condicional son comunes para ajustar las estimaciones en presencia de este tipo de censura, asegurando que las inferencias sobre la supervivencia o la duración de los eventos sean lo más precisas posible (Hosmer et al., 2008).

De forma general, toda variable censurada por la izquierda, se define como:

$$Y_i = \begin{cases} k_i, & \text{si } y_i \leq k_i \\ y_i, & \text{si } y_i > k_i \end{cases}$$

Note que  $P(Y_i = k_i) = P(Y_i \leq k_i) \geq 0$ , ya que  $Y \in \mathbb{R}$ . Luego,  $Y_i = \text{máx}\{y_i, k_i\}, \forall i$ .

### c) Censura por intervalo

La censura por intervalos ocurre cuando se conoce que un evento de interés ha tenido lugar dentro de un rango específico de tiempo, pero no se puede determinar el momento preciso de su ocurrencia dentro de ese rango. Esta forma de censura es particularmente prevalente en estudios que realizan seguimientos en intervalos fijos, como en investigaciones clínicas o estudios de fiabilidad de productos, donde los eventos pueden ser revisados sólo en visitas periódicas o evaluaciones programadas (Kleinbaum y Klein, 2012).

En estos casos, el investigador sabe que el evento ocurrió entre dos observaciones consecutivas; sin embargo, el tiempo exacto del evento sigue siendo incierto. Este tipo de censura introduce una complejidad adicional en el análisis estadístico, ya que los métodos tradicionales que suponen datos precisos sobre los tiempos de eventos podrían llevar a conclusiones erróneas. Se requieren técnicas especializadas que puedan manejar esta incertidumbre inherente a los datos censurados por intervalo (Cleves et al., 2016)

Métodos estadísticos avanzados, incluyendo modelos de riesgo proporcional y técnicas de estimación que utilizan la función de verosimilitud, son esenciales para tratar adecuadamente la censura por intervalo. Estos métodos permiten hacer inferencias sobre la distribución de los tiempos de los eventos, considerando las limitaciones de los datos. Algunas técnicas comúnmente utilizadas involucran el uso de la distribución exponencial y de Weibull para modelar los tiempos de falla, proporcionando un marco más flexible y adecuado para datos censurados de esta naturaleza (Hosmer et al., 2008).

#### 2.1.4. Probabilidades de supervivencia

La probabilidad de supervivencia es una medida estadística crucial en el análisis de supervivencia que describe la proporción de individuos en una población que se espera que sobreviva más allá de un tiempo específico, sin experimentar el evento de interés. Esta probabilidad se calcula utilizando datos de tiempos de supervivencia de una muestra, que pueden incluir datos censurados, proporcionando una visión general de la durabilidad o la longevidad en contextos médicos, técnicos o sociológicos (Arteaga & Moreno, 2020)

### **Métodos de Análisis en Supervivencia**

**1. Métodos Paramétricos:** Los métodos paramétricos implican la suposición de una distribución probabilística específica de los tiempos de supervivencia. Estos métodos son útiles cuando se tiene conocimiento previo de que los datos siguen una distribución como la Exponencial, Weibull, o Log-normal. Estas suposiciones permiten realizar análisis más detallados y predicciones sobre la supervivencia, y son particularmente efectivos cuando la forma de la distribución de supervivencia es bien comprendida a priori (Rossello y González, 2022).

**2. Métodos Semiparamétricos y No Paramétricos:** Estos enfoques no requieren la especificación de una distribución subyacente de tiempos de supervivencia. El método de Kaplan-Meier, por ejemplo, es una técnica no paramétrica que estima la función de supervivencia sin hacer suposiciones sobre la forma de la distribución de los datos. Por su parte, el modelo de riesgos proporcionales de Cox es un enfoque semiparamétrico que permite estimar el efecto de las covariables sobre el riesgo de ocurrencia del evento sin especificar la forma funcional completa de la función de riesgo (Rossello y González, 2022).

### **Aplicaciones de las Probabilidades de Supervivencia**

Las probabilidades de supervivencia son aplicadas en una variedad de contextos para evaluar la efectividad de tratamientos médicos, la fiabilidad de sistemas mecánicos, o la

duración de empleo en estudios laborales, entre otros. Al evaluar, por ejemplo, la efectividad de un nuevo tratamiento médico, las probabilidades de supervivencia calculadas permiten a los investigadores y médicos entender mejor los beneficios potenciales del tratamiento en términos de prolongación de la vida (Felipe et al., 2021)

### ***2.1.5. Funciones básicas de supervivencia***

El análisis de supervivencia, una disciplina estadística, se dedica a explorar los periodos hasta que suceden ciertos eventos significativos. Originalmente concebida para estudiar la duración hasta eventos críticos en pacientes dentro de contextos clínicos, esta metodología ha trascendido hacia sectores tan variados como la ingeniería, las ciencias sociales y las ciencias económicas. Este enfoque se utiliza para analizar no solo la supervivencia humana, sino también la durabilidad de productos o la permanencia de individuos en estudios longitudinales en diversas áreas (Horbath, 2022).

El marco del análisis de supervivencia abarca diversas funciones estadísticas clave. Entre ellas, la función de supervivencia, que estima la probabilidad de que el evento no haya ocurrido en un tiempo  $t$ , y la función de riesgo, que evalúa la probabilidad instantánea de ocurrencia del evento dado que aún no ha ocurrido hasta ese momento. Adicionalmente, se cuenta con la función de riesgo acumulado, que ofrece una acumulación del riesgo a lo largo del tiempo, proporcionando así una visión integral del riesgo al que están expuestos los sujetos del estudio durante el periodo observado (Campos, 2021).

#### **2.1.5.1. Función de supervivencia**

La función de supervivencia, simbolizada como  $S(t)$ , es crucial en el análisis de supervivencia. Esta función matemática describe la probabilidad de que un sujeto o una unidad mecánica sobreviva más allá de un tiempo específico  $t$ , sin que ocurra un evento particular como la muerte o el fallo de un componente. Esta es una herramienta esencial

para los investigadores, ya que proporciona información sobre la durabilidad o el riesgo a lo largo del tiempo (Ramírez et al., 2021).

La función de supervivencia se define formalmente como el complemento de la función de distribución acumulada  $F(t)$ , que es la probabilidad de que el evento haya ocurrido antes o en el tiempo  $t$ . Por lo tanto, la función de supervivencia se calcula como:

$$S(t) = 1 - F(t) \quad (1)$$

y refleja la probabilidad de que el tiempo hasta el evento sea mayor que  $t$  (Guillermo et al., 2022). Se expresa a través de la integral de la función de densidad de probabilidad  $f(t)$  del tiempo hasta el evento, extendiéndose desde  $t$  hasta el infinito:

$$S(t) = \int_t^{\infty} f(t)dt$$

Si "t" es discreta y toma los valores  $t_1 \leq t_2 \leq \dots$ , la función de supervivencia y la función de distribución son calculadas por adición de la función de masa de probabilidad  $p(t_j) = P(T = t_j)$ :

$$S(t) = \sum_{t_j > t} p(t_j)$$

Esta función es inherentemente no creciente, iniciando en  $S(0) = 1$ , indicando que, al inicio, todos los sujetos están vivos o sin fallo, y decrece hacia 0 conforme  $t$  aumenta hasta el infinito, lo que sugiere que la probabilidad de supervivencia se reduce con el tiempo (Alvarado et al., 2020).

La función de supervivencia proporciona una base para la comprensión y modelado estadístico del tiempo de vida, siendo aplicable en una amplia gama de disciplinas, desde estudios médicos hasta ingeniería y ciencias sociales.

### 2.1.5.2. Función de riesgo

La función de riesgo, a menudo denominada tasa de fallo o tasa de riesgo, mide la probabilidad instantánea de que se produzca un evento dado, asumiendo que no ha ocurrido antes de un momento específico  $t$ . Esta función se formula matemáticamente como la derivada del logaritmo negativo de la función de supervivencia en el tiempo  $t$ , lo cual proporciona una estimación de cómo el riesgo de experimentar el evento aumenta o disminuye con el tiempo (Cleves et al., 2016).

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T > t)}{\Delta t} \quad (2)$$

Esta función es valiosa para analizar cómo varía el riesgo de que ocurra el evento a medida que transcurre el tiempo.

Aplicando la probabilidad condicional a la expresión (2) se tiene:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \frac{P(t \leq T < t + \Delta t)}{P(T > t)} \quad (3)$$

Desarrollando y del hecho de que  $P(T > t) = S(t)$  en la expresión (3), se obtiene que:

$$h(t) = \frac{1}{S(t)} \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \quad (4)$$

ahora aplicamos el concepto de la derivada de la ecuación (4), obteniendo lo siguiente:

$$h(t) = \frac{f(t)}{S(t)} \quad (5)$$

realizando la integración de la ecuación (5) se tiene la función de riesgo acumulada.

$$H(t) = \int_0^t h(t) dt = \int_0^t \frac{f(t)}{S(t)} dt \quad (6)$$

a su vez se puede escribir como:  $H(t) = \int_0^t \frac{f(t)}{1-F(t)} dt$  sustituyendo  $u = 1 - F(t)$ , y  $du = -f(t)dt$ , la expresión se convierte en  $H(t) = \int_0^t \frac{-du}{u}$ , de donde resulta que:

$$H(t) = -\ln S(t)$$

$$S(t) = e^{-H(t)}$$

$$h(t) = -\frac{d}{dt}(\ln S(t))$$

$$f(t) = -\frac{d}{dt}S(t)$$

### 2.1.5.3. Función de densidad de probabilidad

La función de densidad de probabilidad, representada por  $f(t)$ , indica la probabilidad de que el evento de interés se presente en un momento específico  $t$ . Esta función es crucial para el análisis detallado de la ocurrencia de eventos en estudios de supervivencia. Está intrínsecamente vinculada a otras funciones estadísticas fundamentales en el análisis de supervivencia, expresándose como  $f(t) = \lambda(t)S(t)$ , donde  $\lambda(t)$  es la función de riesgo y  $S(t)$  la función de supervivencia (Neis et al., 2021). Esto implica que la probabilidad de que el evento suceda en un instante dado es el producto de la tasa de riesgo en ese momento y la probabilidad de que el evento no haya ocurrido antes de ese tiempo.

### 2.1.5.4. Función de riesgo acumulado

La función de riesgo acumulado, denotada como  $H(t)$  se calcula mediante la integración de la función de riesgo,  $h(u)$ , desde el inicio del periodo de observación hasta un tiempo  $t$ . Esta función se define matemáticamente como:

$$H(t) = \int_0^t h(u) du = -\log R(t)$$

donde  $R(t)$  es la función de supervivencia residual, indicando la probabilidad de que el evento no haya ocurrido hasta después del tiempo  $t$  (Rossello & González, 2022). La función de riesgo acumulado ofrece una medida integral del riesgo al que ha estado expuesto un individuo hasta un punto específico en el tiempo, proporcionando una perspectiva acumulativa del riesgo que es esencial para la evaluación de la exposición al riesgo a lo largo del tiempo en estudios longitudinales.

### 2.1.6. *Estimador Kaplan Meier*

El estimador Kaplan-Meier es un método no paramétrico crucial para estimar la función de supervivencia a partir de datos de tiempo hasta evento, especialmente útil en presencia de censura. Este enfoque permite calcular la probabilidad de que un evento no ocurra hasta después de un tiempo específico, proporcionando una visión clara de la distribución del tiempo de supervivencia en un estudio (Monge & Cobo, 2023).

Este estimador se calcula utilizando los tiempos de evento y censura de cada individuo en el estudio. En cada tiempo de evento  $t$ , el estimador de Kaplan-Meier actualiza la probabilidad de supervivencia multiplicando la probabilidad de supervivencia anterior por la proporción de individuos que sobrevivieron más allá de  $t$  sin experimentar el evento. Matemáticamente, si tenemos  $n$  individuos y eventos de interés ocurren en los tiempos  $t_1, \dots, t_k$  con  $d_i$  eventos ocurriendo en cada tiempo  $t_i$  y  $n_i$  sujetos en riesgo justo antes de  $t_i$ , el estimador de la función de supervivencia en cada punto  $t$  se expresa como:

$$\hat{S}(t) = \prod_{j:t_j < t} \frac{n_j - d_j}{n_j}$$

donde  $d_i$  es el número de eventos (fallos) que ocurren en  $t_i$  y  $n_i$  es el número de individuos en riesgo de experimentar el evento justo antes de  $t_i$  (Horbath, 2022). Esta fórmula proporciona una estimación paso a paso de la función de supervivencia, reduciendo su valor cada vez que un evento ocurre y ajustándose en cada paso por la cantidad de sujetos aún en riesgo.

El valor de la función de supervivencia estimada por Kaplan-Meier es particularmente útil porque ofrece una manera intuitiva y gráfica de representar la supervivencia a lo largo del tiempo. Además, permite manejar adecuadamente la censura, asignando un peso completo al tiempo de supervivencia de los individuos hasta que son censurados o alcanzan el evento de interés (Favata et.al. 2023).

### **2.1.7. Estimador de Nelson Aalen**

El estimador de Nelson-Aalen es una herramienta no paramétrica fundamental para estimar la función de riesgo acumulado en estudios de supervivencia. A diferencia del enfoque de Kaplan-Meier, que proporciona estimaciones de la función de supervivencia, el estimador de Nelson-Aalen se centra en acumular la magnitud del riesgo a lo largo del tiempo (Ramírez, et.al, 2021).

Este método estima la función de riesgo acumulado sumando proporciones de fallo a lo largo del tiempo. Para cada intervalo de tiempo donde ocurren uno o más eventos, el riesgo acumulado se incrementa por la proporción de fallos entre los sujetos que estaban en riesgo en ese momento. Matemáticamente, si  $d_j$  es el número de eventos en el tiempo  $t_j$  y  $n_j$  es el número de individuos en riesgo justo antes de  $t_j$  (Ramírez, et.al, 2021), el estimador se calcula como:

$$\hat{H}(t) = \sum_{t_j \leq t} \frac{d_j}{n_j}$$

Donde  $\hat{H}(t)$  es la estimación de la función de riesgo acumulado hasta el tiempo  $t$ . Esta fórmula proporciona una visión integral del riesgo que los individuos han enfrentado hasta cualquier punto en el tiempo, haciéndola particularmente útil para evaluar el riesgo general en estudios longitudinales donde la censura y los tiempos de fallo varían entre los sujetos (Klein et.al, 2014).

La función de Nelson-Aalen es especialmente valiosa en escenarios donde las suposiciones de modelos paramétricos no son adecuadas y donde la distribución exacta de los tiempos de fallo no es conocida.

### **2.1.8. Prueba de hipótesis para la función de supervivencia**

En el campo del análisis de supervivencia, se aplican diversas pruebas estadísticas para analizar y comparar las funciones de supervivencia entre distintos grupos

demográficos o para examinar el efecto de ciertas variables sobre las tasas de supervivencia. Una prueba habitual en estos análisis es la prueba de log-rank, que es esencial para comparar la supervivencia entre múltiples cohortes. Además, otras pruebas como la prueba de Wilcoxon, también conocida como la prueba de Breslow, y la prueba de Mantel-Haenszel también se utilizan frecuentemente (Ovalle, 2022). Estas herramientas estadísticas son cruciales para identificar diferencias estadísticamente significativas en las funciones de supervivencia entre diferentes grupos o condiciones, facilitando así una comprensión más profunda de los factores que pueden afectar la supervivencia en estudios clínicos y de investigación.

$$H_0: S_1(t) = S_2(t) = \dots = S_M(t)$$

$$H_a: S_1(t) \neq S_2(t) \neq \dots \neq S_M(t)$$

Este se hace una contrastación con el p-valor para poder concluir si hay una diferencia significativa o si no exista tal diferencia

### **2.1.8.1. Prueba de Log Rank**

La prueba de Log Rank es una herramienta estadística no paramétrica ampliamente empleada para evaluar diferencias en las funciones de supervivencia entre varios grupos. Este método se centra en determinar si las diferencias observadas en las tasas de supervivencia o incidencia de eventos a lo largo del tiempo entre los grupos son estadísticamente significativas. Funciona mediante la comparación del número de eventos observados con el número de eventos esperados en cada grupo, ajustando por el número de individuos en riesgo en cada intervalo de tiempo (Álvarez et al., 2022).

La hipótesis nula en la prueba de Log Rank postula que no existen diferencias significativas en las funciones de supervivencia entre los grupos estudiados. Un resultado donde el valor p es inferior a 0.05 generalmente indica suficiente evidencia para rechazar la hipótesis nula, sugiriendo que las tasas de supervivencia varían significativamente entre

los grupos. Este método es particularmente valorado en estudios de supervivencia por su eficacia en manejar la censura y por su capacidad para proporcionar resultados claros y directos sobre la igualdad de las curvas de supervivencia en diferentes cohortes (Zuluaga et al., 2022).

La prueba de hipótesis para la prueba de Long Rank es:

$$H_0: S_1(t) = S_2(t) \forall t > 0$$

$$H_a: S_1(t) \neq S_2(t) \forall t > 0$$

El valor estadístico en la prueba de Log Rank se calcula acumulando las discrepancias entre los eventos observados y los eventos esperados en cada punto temporal, y se normaliza por la varianza de estos eventos. En términos prácticos, para un grupo determinado como el grupo 1, el cálculo implica sumar las diferencias entre las muertes observadas y las esperadas, luego dividir esta suma por la varianza de las muertes observadas en ese grupo específico (Álvarez et al., 2022). Esta fórmula permite evaluar si las diferencias observadas son significativas estadísticamente, considerando la fluctuación natural de los datos.

$$\hat{e}_{1i} = \frac{n_{1i}d_i}{n_i}$$

Y la varianza suponiendo una distribución hipergeométrica:

$$\hat{v}_{1i} = \frac{n_{1i}n_{0i}d_i(n_i - d_i)}{n_i^2(n_i - 1)}$$

La estadística de Log Rank para probar la hipótesis de igualdad entre las dos funciones de supervivencia está dado por:

$$Q = \frac{\sum_{i=1}^k (d_{1i} - \hat{e}_{1i})}{\sum_{i=1}^k \hat{v}_{1i}}$$

Donde Q se distribuye como una chi-cuadrada con un grado de libertad ( $p = P(x^2(1)) \geq Q$ ).

Cuando se comparan más de dos grupos usando la prueba de Log Rank, el enfoque se extiende de manera similar al empleado para dos grupos. La metodología subyacente se mantiene constante, aplicando el mismo principio de comparación de eventos observados y esperados entre los grupos. Para facilitar este análisis, se emplea una tabla que organiza los datos de supervivencia de cada grupo, permitiendo el cálculo de los estadísticos de prueba necesarios para evaluar la igualdad entre las múltiples funciones de supervivencia (Rubio y Martínez, 2016). Esta tabla recopila y sistematiza los eventos observados, los esperados, y las varianzas asociadas para cada grupo y tiempo observado, proporcionando una base clara para las comparaciones estadísticas.

**Tabla 1**

*Prueba de igualdad de la función de supervivencia de M grupos de observación ti*

• Evento	• 1	• 2	• ...	• m	• ...	• M	• Total
• Muerte	• $d_{1i}$	• $d_{2i}$	• ...	• $d_{mi}$	• ...	• $d_{Mi}$	• $d_i$
• No muerte	• $n_{1i} - d_{1i}$	• $n_{2i} - d_{2i}$	• ...	• $n_{mi} - d_{mi}$	• ...	• $n_{Mi} - d_{Mi}$	• $n_i - d_i$
• En riesgo	• $n_{1i}$	• $n_{2i}$	• ...	• $n_{mi}$	• ...	• $n_{Mi}$	• $n_i$

Primero se debe estimar el número esperado de muertes de cada grupo bajo la suposición de igual función de supervivencia, es decir:

$$\hat{e}_{1i} = \frac{n_{ki}d_i}{n_i}, i = 1, \dots, m$$

Cuando se analizan múltiples grupos en un estudio de supervivencia, en lugar de calcular una simple varianza, se requiere determinar una matriz de covarianzas asociada al vector de eventos observados  $d_i$ . Esta matriz es crucial para entender las interrelaciones

entre los distintos grupos. La construcción de esta matriz asume que el número de eventos observados en cada grupo sigue una distribución hipergeométrica multivariada (Alvarez, 2023). Los elementos de la diagonal de esta matriz, representados como  $\hat{v}_i$ , se obtienen mediante un método específico que refleja la variabilidad y la dependencia entre los eventos en los distintos grupos analizados.

$$\hat{v}_{mmi} = \frac{n_{mi}(n_i - n_{mi})d_i(n_i - d_i)}{n_i^2(n_i - 1)}, \quad m = 1, 2, \dots, M - 1$$

Los componentes fuera de la diagonal se calculan como:

$$\hat{v}_{lmi} = \frac{n_{li}n_{mi}d_i(n_i - d_i)}{n_i^2(n_i - 1)}, \quad l, m = 1, 2, \dots, M - 1, l \neq m$$

Con los estimadores del número de muertes y la varianza se construye el siguiente estadístico:

$$Q = \left[ \sum_{i=1}^k W_i(d_i - \hat{e}_i) \right]' \left[ \sum_{i=1}^k W_i \hat{V}_i W_i \right]^{-1} \left[ \sum_{i=1}^k W_i(d_i - \hat{e}_i) \right]$$

En este contexto  $W_i$  representa una matriz diagonal cuyos elementos,  $w_i$ , son los pesos correspondientes, y  $d_i$  es el vector que indica el número de eventos, como muertes, para cada grupo en el tiempo  $t_i$ . Por otro lado,  $\hat{e}_i$  es el vector que refleja el número esperado de eventos para cada grupo en el mismo instante (Álvarez et al., 2022). Al igual que en análisis con menos grupos, cuando se examinan múltiples grupos bajo la hipótesis nula que postula igualdad entre las funciones de supervivencia de estos grupos, la estadística  $Q$  resultante sigue una distribución Chi-cuadrado con  $M-1$  grados de libertad, donde  $M$  es el número de grupos ( $p = P(\chi^2(1)) \geq Q$ )

### 2.1.9. Modelo de Cox

El modelo de Cox, también reconocido como el modelo de riesgos proporcionales de Cox, es una técnica estadística prevalente en el análisis de datos de supervivencia. Este método se utiliza ampliamente para investigar la relación entre el tiempo hasta la

ocurrencia de un evento y una o más variables explicativas. El modelo no presupone una forma específica para la función de base de la tasa de riesgo, sino que modela el logaritmo de la tasa de riesgo como una función lineal de las variables explicativas, permitiendo que la tasa de riesgo varíe con el tiempo, pero manteniendo constante el efecto proporcional de las covariables (Torreglosa et al., 2022).

El modelo se expresa con la siguiente fórmula matemática general:

$$h(t, X) = h_0(t) \exp(X\beta^T)$$

donde  $h(t, X)$  es la tasa de riesgo en el tiempo  $t$  para un individuo con un vector de covariables  $X$ ,  $h_0(t)$  es la función de riesgo base, que es común a todos los sujetos, y  $\beta$  es el vector de coeficientes que estima el efecto de cada covariable. Este modelo es particularmente útil por su flexibilidad y la capacidad de incluir diferentes tipos de datos y relaciones complejas entre las variables sin necesidad de especificar la distribución del tiempo de supervivencia (Ortega et al., 2023).

Los coeficientes  $\beta$  en el modelo de Cox se utilizan para calcular el *hazard ratio* (HR), que indica cuánto afecta una unidad de incremento en la variable predictora al riesgo de ocurrencia del evento. Un HR mayor que uno sugiere un aumento en el riesgo asociado con el incremento de la variable, mientras que un HR menor que uno indica una reducción en el riesgo (Cocoletzi et al., 2020).

El modelo de Cox es esencialmente valioso en campos como la medicina y la epidemiología, donde se necesita comprender los factores de riesgo asociados con ciertas condiciones o tratamientos. Además, su capacidad para manejar la censura de datos y su robustez en estudios con múltiples covariables lo hacen indispensable para análisis estadísticos complejos en investigaciones longitudinales.

### 2.1.9.1. Componentes

El modelo de Cox, formalmente conocido como el modelo de regresión de riesgos proporcionales de Cox, es una herramienta estadística avanzada empleada principalmente para estudiar la supervivencia y examinar cómo diversas variables explicativas influyen el riesgo de un evento determinado a lo largo del tiempo. A continuación, se describen detalladamente los componentes esenciales del modelo:

#### a) Variables Predictoras (Covariables)

Las variables predictoras en el modelo de Cox son aquellos factores que se consideran influyentes sobre el riesgo de ocurrencia del evento de interés. Estas variables pueden ser tanto continuas (como la edad o niveles de ciertos marcadores biológicos) como categóricas (como el género o categorías de tratamiento). La influencia de cada covariable se evalúa para determinar cómo afecta el riesgo relativo de experimentar el evento en cuestión (Gómez y Cadarso, 2017)..

#### b) Función de Riesgo Basal (Baseline Hazard Function)

La función de riesgo basal, denotada habitualmente por  $h_0(t)$ , es la tasa de riesgo que se observaría en los individuos si todas las covariables fueran cero. Esta función representa la tasa de ocurrencia del evento de interés en un escenario sin influencias externas de las variables predictoras (Gómez y Cadarso, 2017).. Es una componente clave que se multiplica por el efecto exponencial de las covariables para modelar el riesgo específico de un individuo en cualquier momento dado:

$$hazard_t = \lambda_t = \frac{\text{eventos en el instante } t}{\text{sujetos en riesgo en el instante } t} = \frac{d_t}{n_t}$$

#### c) Hazard Ratio

El hazard ratio (HR) es el coeficiente central en el modelo de Cox y cuantifica el cambio relativo en el riesgo de que ocurra el evento por cada unidad de cambio en la covariable, manteniendo constantes todas las demás variables. Matemáticamente, si

consideramos una covariable  $x$ , el HR para esta covariable se calcularía como  $\exp(\beta_x)$ , donde  $\beta_x$  es el coeficiente estimado para  $x$  en el modelo (Gómez y Cadarso, 2017). Este ratio ofrece una interpretación directa del impacto de las covariables sobre el riesgo:

$$\text{Hazard Ratio (HR)} = \exp(\beta)$$

donde  $\beta$  representa el logaritmo del cambio relativo en el riesgo asociado con las variables predictoras.

### 2.1.9.2. Estimación del modelo de Cox

La estimación de los parámetros en el modelo de regresión de Cox se lleva a cabo mediante el método de la verosimilitud parcial, un procedimiento estadístico avanzado que permite deducir los coeficientes que mejor describen la relación entre las covariables y el riesgo de eventos. Este método se concentra únicamente en los tiempos de los eventos observados, ignorando el tiempo total de seguimiento, lo cual es especialmente útil en presencia de datos censurados.

En el modelo de Cox, los coeficientes ( $\beta$ ) describen la influencia logarítmica de cada covariable sobre la tasa de riesgo, o *hazard*. La función de riesgo en cualquier tiempo  $t$  para un individuo dado, dado un vector de covariables  $X$ , es modelada como:  $\lambda(t|x_1, x_2, \dots, x_k) = \lambda_0(t)e^{\beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k}$  donde  $\lambda_0(t)$  es la función de riesgo basal, que es común a todos los individuos, pero modulada por el efecto exponencial de las covariables.

La razón de riesgo o *hazard ratio* (HR) para una covariable puede ser expresada a través de la exponenciación de los coeficientes del modelo:  $\text{HR} = e^\beta$ . Esto indica cuánto se incrementa o disminuye el riesgo de ocurrencia del evento por cada unidad de incremento en la covariable, manteniendo todas las otras covariables constantes.

La efectividad del modelo de Cox y la interpretación de sus parámetros dependen críticamente de la precisión con la que estos coeficientes son estimados. Los valores de HR superiores a 1 indican un aumento en el riesgo asociado con la covariable, mientras que valores inferiores a 1 indican una reducción en el riesgo.

### 2.1.9.3. Estimación de los parámetros en el modelo de Cox

Los parámetros  $\beta$  del modelo de Cox se estiman maximizando la expresión  $l(\beta) = \sum_{i=1}^D [\beta^t X_i - \log (\sum_{j \in R(t_i)} \exp(\beta^t X_j))]$  que es la log-verosimilitud parcial,  $l(\beta)$ , se adquiere tomando el logaritmo natural de la verosimilitud parcial, esto se realiza resolviendo el sistema de ecuaciones:

$$\frac{\partial l(\beta)}{\partial \beta} = 0$$

Las derivadas parciales de la log-verosimilitud parcial con respecto a  $\beta$  se utilizan para hallar los estimadores de máxima verosimilitud de  $\beta$ .

#### 2.1.9.3.1. Estimación de $\beta$

El modelo de Cox se caracteriza por los coeficientes  $\beta$ , los cuales se estiman a partir de los datos observados. Debido a la presencia de un componente no paramétrico, no es posible aplicar el método completo de máxima verosimilitud. En su lugar, se utiliza la verosimilitud parcial  $L(\beta)$ , una metodología que considera únicamente las probabilidades de los eventos de fallecimiento o falla, dejando de lado los casos censurados. La función de verosimilitud parcial,  $L(\beta)$ , se construye con base en una muestra de  $n$  pacientes, en la que ocurren  $k$  eventos de falla distintos (*donde*  $k \leq n$ ) en los tiempos  $t_1, t_2, \dots, t_k$ . Por lo tanto:

$$\begin{aligned}
& P(\text{desiste (falla en } t_i) / \text{del grupo falla en } t_i) \\
&= \frac{P(\text{desiste (falla en } t_i) / \text{estaba matriculado antes)}}{P(\text{un fallo en } t_i \text{ del grupo en riesgo} / \text{estaba matriculado antes})} \\
&= \frac{h\left(\frac{t_i}{X_i}\right)}{\sum_{j \in R(t_i)} h\left(\frac{t_i}{X_j}\right)} \\
&= \frac{h_0(t_i) \exp(X_i \beta^T)}{\sum_{j \in R(t_i)} h_0(t_i) \exp(X_j \beta^T)} \\
&= \frac{\exp(X_i \beta^T)}{\sum_{j \in R(t_i)} \exp(X_j \beta^T)}
\end{aligned}$$

donde  $R(t_i)$  es el conglomerado de pacientes en riesgo antes de  $t_i$ .

En este sentido, la función de verosimilitud para  $\beta$  se calcula como el producto de varias probabilidades, considerando las fallas  $t_i, i = 1, 2, \dots, n$ .

Esto es;

$$L(\beta) = \prod_{i=1}^n \left[ \frac{\exp(X_i \beta^T)}{\sum_{j \in R(t_i)} \exp(X_j \beta^T)} \right] \delta_1 \quad (7)$$

donde:

$$\delta_1 = \begin{cases} 1, & \text{si } t_i \text{ no es censurado } i = 1, 2, \dots, n \\ 0, & \text{si } t_i \text{ escensurado } i = 1, 2, \dots, n \end{cases}$$

luego aplicando logaritmo a los lados de la expresión (7), se tendrá

$$\begin{aligned}
\ln[L(\beta)] &= \sum_{i=1}^n \delta_1 \ln \left[ \frac{\exp(X_i \beta^T)}{\sum_{j \in R(t_i)} \exp(X_j \beta^T)} \right] \\
&= \sum_{i=1}^n \left[ \delta_1 \left( \ln(\exp(X_i \beta^T)) - \ln \sum_{j \in R(t_i)} \exp(X_j \beta^T) \right) \right]
\end{aligned}$$

$$l(\beta) = \ln[L(\beta)] = \sum_{i=1}^n \delta_1 \left[ (X_i \beta^T) - \ln \left[ \sum_{j \in R(t_i)} \exp(X_j \beta^T) \right] \right] \quad (8)$$

diferenciando la ecuación (8) con respecto al vector de parámetros  $\beta^T$  se tiene:

$$\frac{\partial l(\beta)}{\partial \beta_k} = \sum_{i=1}^n \delta_1 \left[ X_{ik} - \frac{\sum_{j \in R(t_i)} X_{ik} \exp(X_j \beta^T)}{\sum_{j \in R(t_i)} \exp(X_j \beta^T)} \right], k = 1, 2, \dots, p$$

haciendo

$$\frac{\partial l(\beta)}{\partial \hat{\beta}_k} = 0$$

se obtiene el sistema de ecuación por ponderaciones:

$$\sum_{i=1}^n \delta_1 \left[ X_{ik} - \frac{\sum_{j \in R(t_i)} X_{ik} \exp(X_j \beta^T)}{\sum_{j \in R(t_i)} \exp(X_j \beta^T)} \right] = 0, k = 1, 2, \dots, p$$

A partir de este conglomerado de ecuaciones se logra calcular los coeficientes estimados de  $\hat{\beta}$ .

### 2.1.9.3.2. Estimación de $h_0(t)$

Supongamos que  $\hat{h}_0(t)$  es el estimador de  $h_0(t)$ , el cual es invariable entre los tiempos de supervivencia no censurados. Primero, organizamos los tiempos de supervivencia no censurados en orden ascendente, tal como:

$$y_1 \leq y_2 \leq \dots \leq y_k$$

y el conglomerado de riesgos  $R_{(j)}$  correspondiente a estos tiempos ordenados, se define como:

$$R_{(j)} = \{ T_i : T_i \geq y_j \}, j = 1, 2, \dots, k$$

Sea  $\hat{h}_{(0)}, \hat{h}_{(1)}, \hat{h}_{(2)}, \dots$  constantes, o sea:

$$\hat{h}_0(t) = \begin{cases} \hat{h}_0, & 0 \leq t \leq y_1 \\ \hat{h}_{(1)}, & y_1 \leq t \leq y_2 \\ \dots & \dots \end{cases}$$

Se desea encontrar  $\hat{h}_0$  entonces, el paciente del conglomerado de riesgos en  $y_2$  se encuentran en el conglomerado  $R_{(2)}$ . Sabiendo que un paciente falla en  $y_2$ , se tiene el dato  $(y_2, R_{(2)})$ , por lo tanto:

$$\begin{aligned}
1 &= \sum_{j \in R_{(2)}} P \left[ \text{el } j - \text{esimo individuo falla en } y(2) / y(2), R_{(2)} \right] \\
&= \sum_{j \in R_{(2)}} (y_3 - y_2) \hat{h}_{(2)} \exp(X_j \beta^T) \\
&= P[t \leq T < t + \Delta(t)/T \geq t] \\
&= (y_3 - y_2) \hat{h}_{(2)} \sum_{j \in R_{(2)}} \exp(X_j \beta^T)
\end{aligned}$$

Así, el riesgo de probabilidad entre  $y_2$  y  $y_3$  es:

$$(y_3 - y_2) \hat{h}_{(2)} = \frac{1}{\sum_{j \in R_{(2)}} \exp(X_j \beta^T)}$$

ahora utilizamos  $\hat{\beta}$  (obtenido a través del método de máxima verosimilitud parcial) para derivar:

$$\hat{h}_{(2)} = \frac{1}{(y_3 - y_2) \sum_{j \in R_{(2)}} \exp(X_j \beta^T)} = \hat{h}_{(0)}(t)$$

De esta manera, se estima el riesgo base  $h_0(t)$ .

Además, este hallazgo permite obtener la estimación de la función de riesgo acumulado  $\hat{H}_{(0)}(t)$ .

#### 2.1.9.4. Comparación de modelos de Cox

Evaluar y seleccionar el modelo de Cox más apropiado para un conjunto de datos implica comparar diferentes modelos para determinar cuál proporciona la mejor interpretación de los datos sin sobreajustar. Para esto, se utilizan comúnmente el criterio de información de Akaike (AIC) y el criterio de información bayesiano (BIC), que son indicadores cruciales de la calidad del ajuste de un modelo, considerando tanto su precisión como su simplicidad.

- a) **Criterio de Información de Akaike (AIC):** Calcula el equilibrio entre la precisión del modelo y su complejidad, penalizando la inclusión de un mayor número de variables. Se formula como:  $AIC = 2k - 2\ln(L)$ , donde  $k$  es el

número de parámetros en el modelo y  $L$  es la verosimilitud máxima alcanzada. Un valor más bajo de AIC indica un modelo con un mejor balance entre complejidad y capacidad de ajuste (Molina, 2015).

- b) **Criterio de Información Bayesiano (BIC):** Similar al AIC, pero introduce una penalización más fuerte para modelos con más parámetros, especialmente en muestras grandes. Se calcula como:  $BIC = \ln(n)k - 2\ln(L)$ , donde  $n$  es el número de observaciones. Al igual que con el AIC, un valor más bajo de BIC señala un modelo preferible (Córdova y Yurevna, 2019).

#### 2.1.9.5. Índice de concordancia

El índice de concordancia, conocido también como C-statistic o índice C, es una herramienta estadística crucial en el análisis de supervivencia para evaluar el rendimiento predictivo de un modelo de riesgos proporcionales de Cox. Este índice mide la habilidad del modelo para distinguir entre los sujetos que probablemente experimentarán un evento de interés y aquellos que no, en varios puntos en el tiempo. La metodología para calcular el índice de concordancia consiste en evaluar la proporción de todos los pares posibles de individuos en los que el sujeto con el mayor riesgo predicho por el modelo es también el que experimenta primero el evento (Harrell et al., 1996).

El rango del índice de concordancia va de 0.5 a 1.0. Un valor de 0.5 sugiere que el modelo no tiene capacidad discriminativa, equivalente a una elección al azar, mientras que un valor de 1.0 representa una discriminación perfecta entre los sujetos que experimentan y los que no experimentan el evento. Se considera que un índice mayor a 0.7 indica un buen nivel de precisión predictiva del modelo (Harrell et al., 1996).

Para calcular el índice de concordancia, se emplean técnicas estadísticas específicas que permiten una comparación efectiva entre las predicciones del modelo y los eventos observados, facilitando así una evaluación precisa de su aplicabilidad y precisión tanto en contextos clínicos como en investigaciones epidemiológicas (Steyerberg, 2019).

#### **2.1.10. Contraste de hipótesis de Cox**

El contraste de hipótesis en el contexto del modelo de Cox es fundamental para determinar la significancia estadística de las covariables incluidas en el modelo. La hipótesis nula generalmente postula que el coeficiente ( $\beta$ ) para una covariable determinada es cero, lo que implicaría que dicha covariable no tiene un efecto significativo sobre la probabilidad de ocurrencia del evento estudiado. Para poner a prueba esta hipótesis, se emplean métodos estadísticos como el test de Wald y la prueba de razón de verosimilitud (Boj del Val, 2023).

Estas pruebas generan un valor p que, si resulta ser inferior al umbral de significancia establecido (usualmente 0.05), permite rechazar la hipótesis nula. Esto indicaría que existe una relación estadísticamente significativa entre la covariable y el riesgo del evento.

#### **2.1.11. Prueba de razón de verosimilitud**

La prueba de razón de verosimilitud es una técnica estadística esencial para comparar la adecuación de dos modelos de Cox: uno más simple y otro más complejo que incluye covariables adicionales. Esta prueba evalúa si la adición de covariables al modelo básico mejora de manera significativa el ajuste del modelo a los datos observados.

La prueba se basa en la comparación de las verosimilitudes logarítmicas de los dos modelos: el modelo restringido (sin las covariables adicionales) y el

modelo completo (con todas las covariables). La fórmula para calcular la razón de verosimilitud es la siguiente:  $LR = 2\{\log(L(\beta_0)) - \log(L(\hat{\beta}))\}$  donde  $\beta_0$  son los coeficientes del modelo más simple y  $\hat{\beta}$  son los coeficientes estimados del modelo completo después de la inclusión de las covariables adicionales (Hosmer et al., 2011).

Esta medida, que sigue una distribución chi-cuadrado bajo la hipótesis nula de que la inclusión de las covariables adicionales no mejora el ajuste del modelo, permite determinar si las diferencias en el ajuste son estadísticamente significativas.

#### **2.1.12. Test de Wald**

El Test de Wald es una técnica estadística ampliamente utilizada en el análisis de modelos de regresión, incluyendo el modelo de riesgos proporcionales de Cox. Este test se aplica para determinar si los coeficientes estimados en un modelo de regresión difieren significativamente de cero, lo cual indicaría un efecto significativo de las covariables sobre la variable de respuesta (Allison, 2014).

La esencia del Test de Wald es comparar el valor estimado de un coeficiente con su error estándar para determinar si el coeficiente es estadísticamente diferente de cero. El cálculo del estadístico Wald se realiza utilizando la siguiente fórmula:

$$W = (\hat{\beta} - \beta_0)' \hat{\Sigma}_{\hat{\beta}}^{-1} (\hat{\beta} - \beta_0)$$

donde:

$\hat{\beta}$  es el vector de coeficientes estimados del modelo

$\beta_0$  es el valor hipotético del coeficiente bajo la hipótesis nula, generalmente cero.

$\hat{\Sigma}_{\beta}^{-1}$  es la matriz de covarianzas estimada de los coeficientes.

### 2.1.12.1. *Tes de puntuación (Score Test)*

El Test de Puntuación, también conocido como Score Test, es un método estadístico utilizado para evaluar la importancia de un conjunto de covariables dentro de un modelo de regresión, como el modelo de riesgos proporcionales de Cox. Este test es particularmente útil cuando los parámetros del modelo no están completamente especificados o cuando se desea una alternativa más computacionalmente eficiente al test de razón de verosimilitud (Klein y Moeschberger, 2013).

El Score Test se basa en la primera derivada de la función de verosimilitud con respecto a los parámetros bajo la hipótesis nula. No requiere que el modelo bajo la hipótesis alternativa esté ajustado completamente, haciendo este test menos intensivo computacionalmente (Ortega et al., 2023). El test se define por la siguiente fórmula:

$$U(\beta) = \sum_{i=1}^n \int_0^{\infty} [Z_i(t) - \bar{Z}(\beta, t)] dN_i(t)$$

donde:

$\bar{Z}(t)$  representa el promedio ponderado de las covariables para los individuos en riesgo en el tiempo  $t$ , calculado como:  $\bar{Z}(\beta, t) = \frac{\sum_j Y_j(t) r_j(t) Z_j(t)}{\sum_i Y_i(t) r_i(t)}$

$Z_i(t)$  es el vector de covariables para el individuo  $i$  en el tiempo  $t$ .

$dN_i(t)$  es el incremento del contador de eventos para el individuo  $i$  en el tiempo  $t$ , indicando si el evento de interés ocurrió.

$Y_i(t)$  es un indicador de si el individuo  $i$  está en riesgo en el tiempo  $t$ .

Este estadístico,  $U(\beta)$ , se evalúa bajo la hipótesis nula de que no hay efecto (todos los coeficientes son cero), y típicamente sigue una distribución chi-cuadrado, permitiendo determinar si las covariables tienen un efecto significativo en la tasa de fallos modelada por el Cox. El uso del Score Test es crucial en situaciones donde la estimación completa de modelos más complejos es inviable o innecesariamente costosa.

## 2.2. Marco conceptual

### 2.1.1. Desistimiento

El desistimiento es entendido como el acto voluntario y unilateral por el cual una parte, ya sea el actor o el demandado, decide poner fin a su participación en un proceso judicial, renunciando de manera total o parcial a sus pretensiones, acciones o derechos, ya sea durante el curso del juicio o antes de que se dicte una resolución definitiva. Este acto debe ser comunicado formalmente y puede requerir aceptación por parte de la otra parte o autorización judicial dependiendo del marco normativo aplicable (Sánchez, 2020, p. 89)

#### **Características principales:**

- a) **Irrevocable:** Una vez realizado el desistimiento, no se puede volver atrás.
- b) **No requiere consentimiento:** La otra parte no tiene que aprobar el desistimiento.
- c) **Efecto de cosa juzgada:** El desistimiento produce el mismo efecto que una sentencia firme, impidiendo que se inicie un nuevo proceso sobre la misma pretensión.

#### **Tipos de desistimiento:**

- a) **Desistimiento de la acción o pretensión:** Se renuncia al derecho material que sustenta la demanda.
- b) **Desistimiento del proceso:** Se pone fin al proceso judicial sin renunciar al derecho material.

### ***2.2.2. Desistimiento Universitario***

En el ámbito universitario, el desistimiento se refiere a la acción por la cual un estudiante decide cancelar su matrícula en una carrera o programa académico. De acuerdo con la Ley Universitaria N° 30220 (Perú, 2014), las universidades deben establecer un plazo específico dentro del cual los estudiantes pueden ejercer su derecho al desistimiento sin penalización académica o económica. Este plazo varía según la institución y el programa de estudios, y debe estar claramente especificado en los reglamentos internos de cada universidad.

#### **Derecho de Desistimiento:**

Las universidades suelen establecer un plazo específico dentro del cual los estudiantes pueden ejercer su derecho al desistimiento sin penalización académica o económica. Este plazo varía según la institución y el programa de estudios.

#### **Consecuencias del Desistimiento:**

- a) **Cancelación de Matrícula:** Se da de baja del programa académico y se pierde la condición de estudiante.
- b) **Devolución de Pagos:** En algunos casos, se puede solicitar la devolución total o parcial de los pagos realizados por concepto de matrícula, dependiendo del plazo y las políticas de la universidad.
- c) **Posibilidad de Reingreso:** La posibilidad de reingresar a la misma u otra carrera en la misma universidad o en otra institución dependerá de las normas y reglamentos de cada institución.

### **Factores a Considerar antes de Desistir:**

- a) **Consecuencias Académicas:** El desistimiento puede afectar el historial académico y las posibilidades de continuar estudios superiores.
- b) **Implicaciones Financieras:** Se deben considerar los costos de matrícula ya pagados y la posibilidad de reembolso.
- c) **Análisis de las Razones:** Es importante reflexionar sobre las causas del desistimiento y evaluar si existen alternativas para solucionarlas.

### **2.2.3. Deserción estudiantil**

La deserción estudiantil es el abandono de los estudios por parte de los estudiantes antes de completar su carrera. Se considera un problema grave en todo el mundo, ya que tiene un impacto negativo no solo en los estudiantes, sino también en las universidades y la sociedad en general.

La Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) define la deserción estudiantil como "el abandono de la escuela por parte de un estudiante matriculado antes de completar el ciclo educativo de que se trate" (UNESCO, 2015).

La UNESCO también distingue entre dos tipos de deserción estudiantil:

- a) **Deserción temprana:** Es el abandono de la escuela antes de completar el nivel primario.
- b) **Deserción tardía:** Es el abandono de la escuela después de completar el nivel primario, pero antes de completar el nivel secundario o terciario.

Es importante señalar que el desistimiento universitario puede considerarse un tipo de deserción tardía, específicamente cuando el estudiante decide cancelar su matrícula en la universidad antes de completar su programa académico. El desistimiento, aunque de carácter voluntario, también tiene

implicaciones sociales y académicas, ya que refleja una interrupción en el proceso educativo del estudiante.

#### **Deserción estudiantil según:**

**Tinto** (1989), afirma que el estudio de la deserción en la educación superior es extremadamente complejo, ya que implica no solo una variedad de perspectivas sino, además, una gama de diferentes tipos de abandono.

**Gonzales** (2005), la deserción tiene consecuencias sociales en términos de las expectativas de los estudiantes y sus familias; emocionales es por la disonancia entre las aspiraciones de los jóvenes y sus logros y, también, importantes consecuencias económicas tanto para las personas como para el sistema en su conjunto.

Braxton et al. (1997), los enfoques del análisis de la deserción pueden ser agrupadas en cinco grandes categorías, dependiendo del énfasis que otorgan las variables explicativas, ya sean individuales, institucionales o del medio familiar. De esta manera es posible reconocer los siguientes enfoques: Psicológico, Económicos, Sociológicos, Organizacionales y de Interacciones

#### ***2.2.4. Factores que contribuyen a la deserción estudiantil***

Los factores que contribuyen a la deserción estudiantil pueden ser de naturaleza económica, académica, personal o institucional.

##### **a) Factores económicos**

Los factores económicos incluyen la falta de recursos económicos para pagar la matrícula, los libros y otros gastos relacionados con la educación. Los estudiantes de familias con bajos ingresos tienen más probabilidades de desertar de la escuela que los estudiantes de familias con ingresos altos.

##### **b) Factores académicos**

Los factores académicos incluyen las dificultades académicas, el fracaso escolar y la falta de motivación para estudiar. Los estudiantes que tienen dificultades académicas tienen más probabilidades de desertar de la escuela que los estudiantes que no tienen dificultades académicas.

**c) Factores personales**

Los factores personales incluyen problemas familiares, de salud o de pareja. Los estudiantes que enfrentan problemas personales tienen más probabilidades de desertar de la escuela que los estudiantes que no enfrentan problemas personales.

**d) Factores institucionales**

Los factores institucionales incluyen la calidad de la educación, la falta de apoyo a los estudiantes y la burocracia. Los estudiantes que asisten a escuelas con una mala calidad de la educación tienen más probabilidades de desertar de la escuela que los estudiantes que asisten a escuelas con una buena calidad de la educación.

**2.2.5. Impacto de la deserción estudiantil**

La deserción estudiantil tiene un impacto negativo en los estudiantes, las universidades y la sociedad en general.

**a) Impacto en los estudiantes**

Los estudiantes que desertan de la escuela tienen menos probabilidades de encontrar un empleo bien remunerado, lo que puede conducir a la pobreza y la exclusión social. Los estudiantes que desertan de la escuela también tienen más probabilidades de involucrarse en conductas delictivas o de caer en la adicción a las drogas o al alcohol.

### **b) Impacto en las universidades**

Las universidades pierden recursos y prestigio cuando los estudiantes desertan. Las universidades también tienen que invertir recursos para atender a los estudiantes que están en riesgo de desertar.

### **c) Impacto en la sociedad**

La sociedad pierde la oportunidad de contar con profesionales bien capacitados cuando los estudiantes desertan de la escuela. La deserción estudiantil también puede conducir a un aumento de la delincuencia y la pobreza.

#### ***2.2.6. Medidas para prevenir la deserción estudiantil***

La identificación de los factores que contribuyen a la deserción estudiantil es un paso importante para prevenirla. Una vez que se conocen los factores que contribuyen a la deserción, se pueden tomar medidas para abordarlos. Estas medidas pueden incluir:

- a) **Programas de apoyo financiero:** Estos programas pueden ayudar a los estudiantes a pagar la matrícula, los libros y otros gastos relacionados con la educación.
- b) **Programas de tutorías:** Estos programas pueden ayudar a los estudiantes con dificultades académicas.
- c) **Programas de orientación vocacional:** Estos programas pueden ayudar a los estudiantes a elegir una carrera que se ajuste a sus intereses y habilidades.
- d) **Reformas curriculares:** Estas reformas pueden mejorar la calidad de la educación y hacerla más relevante para las necesidades de los estudiantes

## **2.3. Antecedentes de estudio**

### ***2.3.1. Antecedentes internacionales***

Paz (2022) utiliza el análisis de supervivencia para investigar la deserción estudiantil en la carrera de Ingeniería Civil en la Universidad Nacional de Tucumán. A través de técnicas estadísticas como el estimador de Kaplan-Meier y el modelo de riesgos proporcionales de Cox, se analizó el tiempo hasta el abandono de los estudiantes, junto con factores como el rendimiento académico en las primeras materias. Los resultados indican que el 41% de los abandonos ocurre durante el primer año, y un 58% de los desertores no aprobaron ninguna materia. Se observa que las asignaturas del primer cuatrimestre, como Cálculo I y Física I, tienen una fuerte influencia en la probabilidad de abandono si no se regularizan a tiempo. Además, el análisis por género mostró que las mujeres tienen un mayor riesgo de deserción que los varones. En conclusión, el estudio destaca la importancia de intervenir en las primeras etapas académicas para reducir la tasa de abandono.

Vélez (2020) llevó a cabo un estudio sobre la deserción estudiantil en Medellín, aplicando el Análisis de Supervivencia y el Modelo de riesgos proporcionales de Cox para determinar los factores que influyen en la deserción de los estudiantes en educación superior. También se utilizó el Análisis de Componentes Principales (ACP) para construir un índice de permanencia estudiantil y prever el riesgo de deserción. La muestra incluyó 518 estudiantes de dos cohortes en programas universitarios y tecnológicos, y se observó que la deserción era más elevada en los primeros semestres. Los modelos estadísticos permitieron identificar los factores sociodemográficos y académicos asociados a la deserción, y el índice de permanencia facilitó la generación de alertas

tempranas, permitiendo intervenir en los casos de mayor riesgo. Este enfoque resultó efectivo para mejorar la retención estudiantil mediante intervenciones preventivas.

Quiñones et al. (2020) desarrolló un estudio cuyo objetivo fue aplicar técnicas de minería de datos para estimar la deserción estudiantil de los pueblos Awajún y Wampis en la Universidad Nacional de Jaén. Se utilizó la metodología CRISP-DM, que incluyó el análisis de datos de 49 estudiantes a través del software Weka. Los resultados mostraron que el 45% de los estudiantes abandonaron sus estudios, influenciados por cinco variables clave: cursos aprobados, créditos aprobados, comunidad de origen, promedio acumulado y ciclo de ingreso. Se presentaron tres modelos predictivos con un 87.8% de precisión, concluyendo que aquellos estudiantes con menos de 10 cursos aprobados o 27 créditos aprobados tienen mayor probabilidad de desertar.

Jiménez (2021) realizó un estudio en el que se analizaron los factores que influyen en la deserción de estudiantes doctorales en la Universidad de los Andes, Colombia. Con un enfoque mixto, empleó un análisis cuantitativo mediante los modelos de supervivencia de Nelson-Aalen y los estimadores de riesgo de Hazard, además de un análisis cualitativo basado en entrevistas a 15 estudiantes y directores de programas. A partir de una muestra de 719 estudiantes, se encontró que la probabilidad de deserción aumentaba hasta un 24% conforme avanzaban los semestres, siendo más alta entre las primeras ocho matrículas. Los factores que contribuyeron a la deserción incluyeron aspectos económicos, académicos, familiares y de salud, tanto mental como física.

Medina y Romero (2020) realizaron un estudio en la URACCAN-Nueva Guinea para analizar los factores asociados al abandono escolar y proponer un

modelo de retención. A través de un enfoque mixto y un análisis correlacional y fenomenológico, emplearon los modelos Kaplan-Meier y regresión de Cox con una muestra de 347 estudiantes. Los resultados indicaron que un 61.7% de los estudiantes abandonaron sus estudios antes de 2017, siendo las carreras con mayor deserción Agroforestal, Informática y Zootecnia. Las variables más influyentes fueron el número de asignaturas reprobadas y el promedio académico. En conclusión, se propuso un plan de retención estudiantil para mitigar la deserción.

### **2.3.2. Antecedentes nacionales**

Saavedra (2023) aplicó un modelo de riesgos competitivos bayesiano para analizar los factores que influyen en la deserción y graduación de estudiantes de la Pontificia Universidad Católica del Perú (PUCP) entre 2004 y 2012. Con datos de 7,002 estudiantes, el estudio mostró que los estudiantes que ingresaron a la universidad por su primera opción tenían una mayor probabilidad de graduarse y una menor probabilidad de abandonar. Comparado con el enfoque frecuentista, el modelo bayesiano resultó más eficiente para predecir estos eventos, lo que destacó su utilidad en situaciones de riesgos competitivos.

Basurto (2021) investigó la relación entre los factores socioeconómicos, académicos e individuales y la deserción estudiantil en una universidad privada de Lima durante el año 2019. Utilizando un enfoque cuantitativo y un diseño descriptivo-correlacional, se aplicó un cuestionario a una muestra de 145 estudiantes desertores. Los resultados revelaron una correlación inversa significativa ( $Rho = -0.669$ ) entre los factores analizados y la deserción, siendo los factores socioeconómicos y académicos los más influyentes en la decisión de abandonar los estudios.

Cerron y Saenz (2023) investigaron los factores de riesgo que contribuyeron a la deserción escolar durante la pandemia de COVID-19 en la Institución Educativa N° 30069 Callaballauri, Chupaca. El estudio, con enfoque cualitativo y diseño fenomenológico, se basó en una muestra de 12 estudiantes de primaria. Los resultados revelaron que los factores económicos, educativos y familiares jugaron un papel clave en el abandono escolar. La falta de acceso a dispositivos tecnológicos y la incapacidad de los padres para apoyar adecuadamente en las tareas escolares fueron las principales causas del abandono educativo durante este período.

Jiménez (2022) realizó un estudio descriptivo sobre la deserción estudiantil en el CETPRO N°005 “El Tablazo” en Corrales, Tumbes, utilizando análisis estadístico basado en la recopilación de datos de los últimos cinco años. Se aplicaron encuestas y técnicas de análisis de experiencias para identificar patrones de deserción. Los resultados mostraron que la falta de recursos económicos y la desmotivación fueron las principales causas del abandono escolar. A partir del análisis estadístico, se propusieron medidas correctivas enfocadas en mejorar la retención estudiantil mediante estrategias de apoyo y motivación.

### ***2.3.3. Antecedentes locales***

Paucar y Quispe (2021) aplicaron un análisis de supervivencia para estudiar los factores asociados a la deserción estudiantil en la Escuela Profesional de Matemática de la Universidad Nacional de San Antonio Abad del Cusco. Utilizando los modelos de Kaplan-Meier y la regresión de Cox, analizaron una cohorte de 131 estudiantes ingresantes entre 2010-I y 2012-II, haciendo un seguimiento durante 10 semestres. Los resultados revelaron una tasa de deserción

del 61.8%, influenciada por factores como el sexo, la edad, el ingreso económico familiar y la ocupación del jefe de hogar. Estos hallazgos proporcionan información útil para la implementación de estrategias de retención.

## CAPÍTULO III

### HIPOTESIS Y VARIABLES

#### 3.1. Hipótesis

##### 3.1.1. *Hipótesis general*

- a) El modelo de riesgos proporcionales de Cox explica de manera significativa el desistimiento estudiantil en la Escuela Profesional de Matemática de la Facultad de Ciencias Químicas, Físicas y Matemática de la UNSAAC durante el periodo 2015-2023.

##### 3.1.2. *Hipótesis específicas*

- a) La probabilidad de que los estudiantes de la Escuela Profesional de Matemática de la Facultad de Ciencias Químicas, Físicas y Matemática de la UNSAAC, continúen sus estudios varía significativamente a lo largo del tiempo durante el periodo 2015-2023.
- b) El riesgo de desistimiento estudiantil se acumula significativamente conforme transcurre el tiempo durante el periodo 2015-2023.
- c) El sexo, el lugar de procedencia y la condición socioeconómica contribuyen significativamente al desistimiento estudiantil en la Escuela Profesional de Matemática de la Facultad de Ciencias Químicas, Físicas y Matemática de la UNSAAC durante el periodo 2015-2023.

#### 3.2. Identificación de variables e indicadores

##### **VARIABLES INDEPENDIENTES**

Las variables independientes en este estudio representan diversas características que se consideran como posibles factores de desistimiento estudiantil en la E.P de matemática. Incluyen la edad, el sexo, el lugar de

procedencia, la condición económica, así como el número de hermanos de los estudiantes matriculados.

**Variables dependientes:**

Tiempo de desistimiento.

Evento: desistimiento de estudiantes matriculados en la E.P de matemática. Esta variable mide si un estudiante matriculado desistió o no durante el periodo de estudio.

### 3.3. Operacionalización de variables

**Tabla 2**

*Matriz de operacionalización de variables*

VARIABLE	DEFINICION CONCEPTUAL	DEFINICION OPERACIONAL	ESCALA	TIPO DE VARIABLE	UNIDAD DE MEDIDA
<b>VARIABLES INDEPENDIENTES</b>					
Edad	La edad de los estudiantes en el momento de la matricula	Registro de la edad en años cumplidos al momento de la matricula.	Razón	Cuantitativa discreta	Años cumplidos.
Sexo	La clasificación de los estudiantes como masculinos o femeninos.	Registro del sexo asignado al estudiante.	Nominal	Cualitativa dicotómica	1: femenino 2: masculino
Condición económica	La clasificación de la condición económica del estudiante	Registro de la condición económica de cada estudiante.	Nominal	Cualitativa dicotómica	0:Pobre 1:No pobre
Lugar de procedencia	El lugar exacto de procedencia de los estudiantes.	Registro del lugar de procedencia de los estudiantes.	Nominal	Cualitativa politómica	

Numero de hermanos	Es la cantidad de hermanos de sangre del estudiante matriculado.	Registro del número exacto de hermanos que tienen los estudiantes	Razón	Cuantitativa discreta
--------------------	--	---	-------	-----------------------

---

**VARIABLES DEPENDIENTES**

---

Tiempo de desistimiento	Tiempo desde la matricula del estudiante hasta el desistimiento	Tiempo transcurrido hasta el desistimiento o la ausencia del mismo entre los estudiantes matriculados.	Razón	Cuantitativa discreta	Periodo semestral
Evento	Desistimiento de los estudiantes matriculados	Mide si un estudiante matriculado desiste o no por diversos motivos.	Nominal	Cualitativa dicotómica	0:no desiste 1: desiste

---

## **CAPÍTULO IV**

### **METODOLOGÍA**

#### **4.1. Ámbito de estudio: Localización política y geográfica**

El ámbito de estudio del presente trabajo de investigación denominado “DESISTIMIENTO ESTUDIANTIL EN LA ESCUELA PROFESIONAL DE MATEMATICA DE LA FACULTAD DE CIENCIAS QUIMICAS, FISICAS Y MATEMATICA-UNSAAC- PERIODO 2015-2023”, en la Escuela Profesional de Matemática de la Facultad de Ciencias Químicas, Físicas y Matemáticas de la Universidad Nacional de San Antonio Abad del Cusco, teniendo como domicilio legal la Av. de La Cultura 773 del Distrito de Cusco, Provincia de Cusco y Región Cusco

#### **4.2. Tipo y nivel de la investigación**

El tipo de investigación que se utilizó fue una investigación aplicada porque según Sánchez y Reyes (2018). Esta modalidad se centra en la resolución de problemas específicos que afectan a individuos y a la sociedad en general. En consecuencia, ofrece la oportunidad de encontrar soluciones tangibles y prácticas para los desafíos identificados.

El nivel al que corresponde la investigación es descriptivo, ya que el estudio busca detallar y caracterizar el estado de los estudiantes matriculados durante su permanencia en términos de supervivencia, describiendo las variables y factores asociados a lo largo del tiempo especificado (Hernández & Mendoza, 2018). Esto implica la recopilación de datos precisos sobre la duración de la permanencia de estudio, tasas de supervivencia y otros aspectos relevantes.

Adicionalmente, como investigación explicativa, ya que tiene como objetivo profundizar más allá de la descripción, intentando identificar y analizar las causas o factores que influyen en los resultados de supervivencia de estos estudiantes.

### **4.3. Método de investigación**

El método efectuado en la indagación fue el hipotético deductivo, tal como lo describen Hernández y Mendoza (2018). Este método se centra en la formulación de hipótesis que posteriormente son sometidas a prueba mediante la recolección y análisis de datos.

### **4.4. Diseño de investigación**

Para Hernández y Mendoza (2018) el diseño no experimental es aquel en el que no se lleva a cabo ninguna manipulación de la variable independiente, y longitudinal porque se examina el fenómeno a lo largo de uno o más periodos de tiempo. En este sentido, el diseño a emplear es de tipo no experimental y longitudinal, ya que se mostrará la evolución de la permanencia de los estudiantes matriculados en la E.P de Matemática en su entorno natural, sin intervención deliberada en los factores relacionados con este evento. El estudio se enfocará en analizar a lo largo del tiempo a los estudiantes matriculados en la E.P de Matemática en el periodo 2015-2023 retrospectivo.

### **4.5.Unidad de Análisis**

La unidad de análisis es cada estudiante matriculado de la Escuela Profesional de Matemática de la Facultad de Ciencias Químicas, Físicas y Matemática-UNSAAC-Periodo 2015-2023.

### **4.6. Población y muestra**

La población es el conjunto de todos los casos que concuerdan con una serie de especificaciones. Lo que lleva al presente estudio a delimitar la población que va a ser estudiada y sobre la cual se pretende generalizar los resultados (Hernández, 2012). En este caso, la población está constituida por la totalidad de estudiantes matriculados en la Escuela Profesional de Matemática de la Facultad de Ciencias Químicas, Físicas y Matemáticas de la UNSAAC durante el periodo 2015-2023. De acuerdo con los registros

obtenidos, se presenta la siguiente información sobre el total de estudiantes matriculados por año: en 2015 se registraron 45 estudiantes, en 2016 fueron 12, en 2017 se matricularon 38, en 2018 se alcanzaron 43, en 2019 se registraron 23, en 2020 se matricularon 19, en 2021 se tuvo un incremento con 62 estudiantes, en 2022 fueron 46 y en 2023 se registraron 38 estudiantes, lo que hace un total de 326 estudiantes matriculados en el periodo de estudio.

El presente estudio ha decidido trabajar con la totalidad de la población para garantizar resultados más confiables y evitar sesgos asociados a la selección de una muestra. Además, esta decisión permite incorporar datos censurados en el análisis, lo que resulta esencial para la aplicación de técnicas de análisis de supervivencia en este estudio. Por tanto, no se considera el uso de muestras ni técnicas de muestreo.

#### **4.7. Técnicas e instrumentos de recolección de datos**

##### **Técnica**

La técnica de captación de datos fue la observación documental de las listas de los estudiantes matriculados por semestre, proporcionada por el centro de cómputo de la UNSAAC, en un período comprendido entre los años 2015 -2023. Con estas listas fue posible, obtener una nueva lista de aquellos estudiantes (población), que desertaron de la Escuela Profesional de Matemáticas en el periodo indicado, que sirvió para hacer los estudios que este trabajo

##### **Instrumento**

El instrumento fue una ficha de registro de los datos de las listas de los estudiantes matriculados por semestre, en un período comprendido entre los años 2015 al 2023 elaborado por el investigador.

#### **4.8. Técnicas de análisis e interpretación de la información**

El análisis descriptivo se realizó mediante tablas de frecuencia, porcentajes, y medidas de tendencia central y dispersión para las variables explicativas. Cada tabla incluyó su respectivo gráfico y una interpretación de los resultados.

En este estudio, se aplicó el análisis de supervivencia para evaluar el tiempo transcurrido hasta el desistimiento de los estudiantes en la carrera. El tiempo de supervivencia se definió como la diferencia entre la fecha de ingreso del estudiante y la fecha en que ocurrió su desistimiento. A los estudiantes que no desistieron, o aquellos cuya situación actual no se pudo determinar, se les asignó un estado 0. También se asignó el estado 0 a los estudiantes que se retiraron antes del período de estudio. En cambio, aquellos que desistieron durante el período de estudio se asignó el estado 1. Los estudiantes con estado 0 fueron considerados censurados, dado que no se pudo determinar si continuaron en la carrera o si desistieron. En el presente estudio se trabajó con datos censurados a la derecha. La censura a la derecha ocurre cuando no se observa el evento de interés (desistimiento del estudiante) durante el período de estudio, ya sea porque el estudiante sigue activo al finalizar el período, abandonó la carrera sin fecha exacta de retiro o porque no se tiene información disponible sobre su situación actual.

Posteriormente, se estimó la función de supervivencia con el método de Kaplan-Meier y la función de riesgo acumulado con Nelson-Aalen. Para comparar las curvas de supervivencia entre categorías, se aplicó la prueba de Log-Rank. Luego, se ajustó el modelo de riesgos proporcionales de Cox, seleccionando el mejor modelo con los criterios AIC y BIC. Todo el análisis se realizó utilizando el software R.

## CAPÍTULO V

### RESULTADOS Y DISCUSIÓN

#### 5.1. Procesamiento, análisis e interpretación

##### 5.1.1 Análisis exploratorio de datos

##### 5.1.1.1. Exploración de datos para la variable sexo

**Tabla 3**

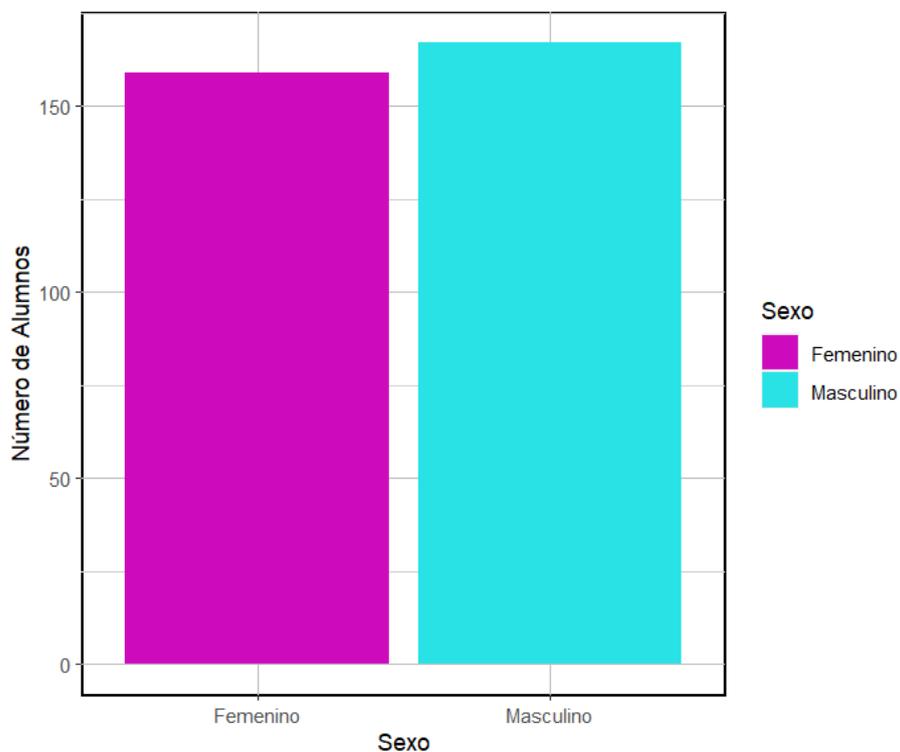
*Sexo de los estudiantes matriculados durante 2015-2023 EP. Matemática UNSAAC*

Sexo	N° de estudiantes	%
Femenino (1)	159	49%
Masculino (2)	167	51%

*Nota.* Los números entre paréntesis indican el código utilizado para representar el sexo de los estudiantes.

**Figura 1**

*Sexo de los estudiantes matriculados durante 2015-2023 EP. Matemática UNSAAC*



La Tabla 3 muestra que son 159 estudiantes de sexo femenino y que representan el 49% del total, mientras que los estudiantes de sexo masculino son 167 y esto representa al 51% del total de estudiantes del estudio.

En la Figura 1 se observa que la barra morada representa el número de estudiantes de sexo femenino, mientras que la barra celeste representa el número de estudiantes de sexo masculino. La Figura 1 muestra que la mayoría de los estudiantes matriculados en la E.P de Matemática de la UNSAAC eran de sexo masculino.

#### 5.1.1.2. Exploración de datos para la variable condición económica

**Tabla 4**

*Condición Económica de los estudiantes matriculados durante 2015-2023 EP.*

*Matemática UNSAAC*

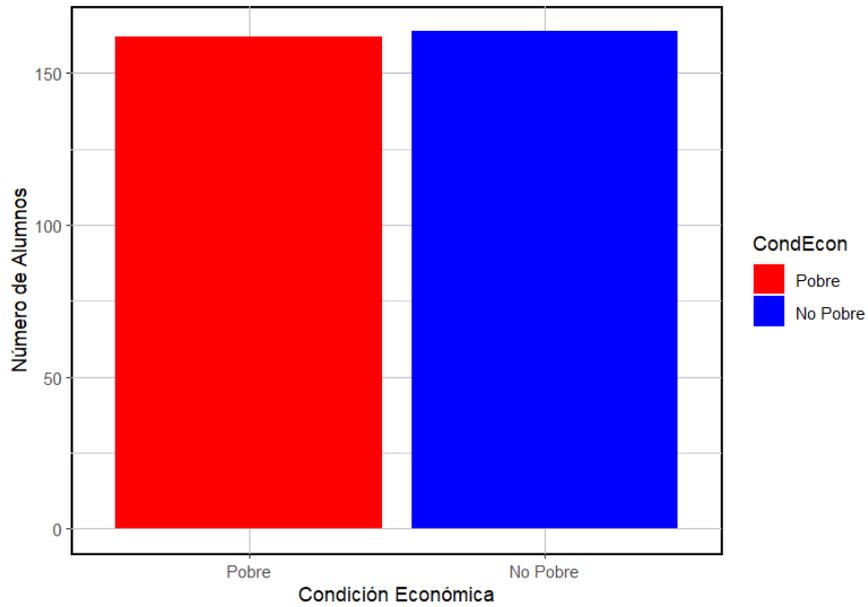
Condición	Nº de estudiantes	%
Pobre (0)	162	49.70%
No pobre (1)	164	50.30%

*Nota.* Los números entre paréntesis indican el código utilizado para representar la condición económica de los estudiantes.

## Figura 2

*Condición Económica de los estudiantes matriculados durante 2015-2023 EP.*

*Matemática UNSAAC*

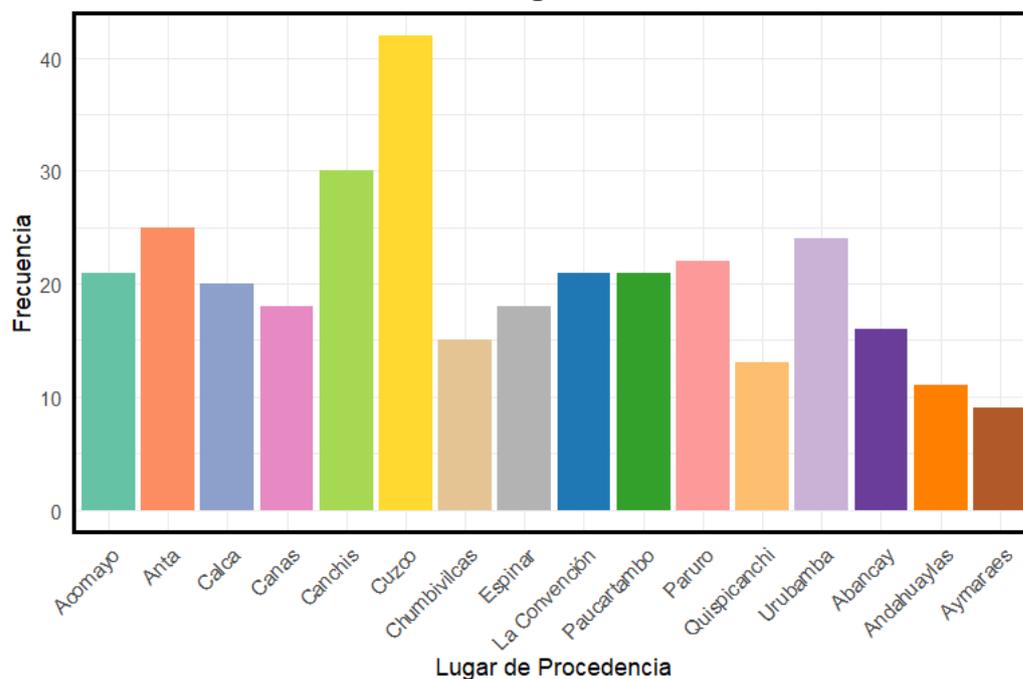


La Figura 2 es un diagrama de barras que representa los datos de la Tabla 4 . La Figura 2 muestra que los estudiantes matriculados durante el periodo de estudio de la E.P de Matemática de la UNSAAC son de condición económica no pobre y pobre no visualiza diferencia.

### 5.1.1.3. Exploración de datos para la variable lugar de procedencia

**Figura 3**

*Lugar de procedencia de los estudiantes matriculados durante el periodo de estudio de la E.P Matemática de la UNSAAC*



En la Figura 3 se observa la distribución del lugar de procedencia de los estudiantes matriculados en la Escuela Profesional de Matemática de la UNSAAC durante el periodo de estudio. Se puede notar que la mayor frecuencia de estudiantes proviene de Cuzco, con un total de aproximadamente 43 estudiantes, lo que representa el valor más alto en comparación con otros lugares de procedencia.

Otros lugares con una alta cantidad de matriculados son Canchis y Acomayo, que cuentan con alrededor de 30 y 25 estudiantes, respectivamente. En contraste, distritos como Andahuaylas y Aymaraes tienen la menor representación, con menos de 10 estudiantes.

#### 5.1.1.4. Exploración de datos para la variable edad

**Tabla 5**

*Estadísticas de la Edad de los estudiantes matriculados durante el periodo de estudio de la E.P Matemática de la UNSAAC*

Edad					
Min	Q1	Mediana	Media	Q3	Max
17	19	22	24	28	35

*Nota.* Min (Edad mínima), Q1: Primer cuartil, Mediana (Edad mediana), Media (Edad promedio), Q3(Tercer cuartil), Max (Edad máxima).

La Tabla 5 presenta estadísticas descriptivas de la edad de los estudiantes matriculados de la E.P Matemática de la UNSAAC durante los años 2015 - 2023. Los valores mínimos y máximos son 17 y 35 años respectivamente. El primer cuartil (Q1) es 19 años, la mediana es 22 años, la media es aproximadamente 23.65 años y el tercer cuartil (Q3) es 28 años. Estos datos indican que la mayoría de los estudiantes matriculados se encuentran entre los 19 y 28 años, con una edad media de alrededor de 24 años.

#### 5.1.1.5. Exploración de datos para la variable número de hermanos

**Tabla 6**

*Numero de hermanos de los estudiantes matriculados durante el periodo de estudio de la E.P Matemática de la UNSAAC*

N° de Hermanos	Cantidad	%
1	49	15.03
2	56	17.18
3	58	17.79
4	65	19.94
5	50	15.34

6                      48                      14.72

---

La Tabla 6 presenta la cantidad de hermanos de los estudiantes matriculados durante el periodo de estudio 2015-2023, se puede apreciar que la mayoría de estudiantes tienen 3 hermanos mientras que la menor cantidad de hermanos es de 1.

**Tabla 7**

*Resumen para la variable evento de los estudiantes matriculados durante el periodo de estudio de la E.P Matemática de la UNSAAC*

---

Total	N°. Eventos	N°. de censurados	Eventos %	Censurados %	Mediana	LI 0.95	LS 0.95
326	298	28	91.41 %	8.6 %	6	6	9

---

En la Tabla 7 se presenta un resumen del desistimiento de los estudiantes matriculados en la Escuela Profesional de Matemática de la UNSAAC durante el periodo de estudio. De un total de 326 estudiantes, 298 experimentaron el evento de desistimiento, mientras que 28 estudiantes fueron censurados, es decir, no desertaron durante el periodo de observación.

La mediana del tiempo hasta el desistimiento fue de 6 periodos, con un intervalo de confianza del 95% entre 6 y 9 periodos, lo que indica que la mitad de los estudiantes desertaron antes de este tiempo. Esto sugiere que el desistimiento es un fenómeno frecuente dentro de la escuela profesional de matemática, con una variación moderada en el tiempo hasta su ocurrencia.

Del 100% de estudiantes matriculados en la E.P Matemática de la UNSAAC durante el periodo 2015-2023, el 91.41 % de estudiantes presentaron el evento de interés y el 8.6 % del resto de los estudiantes fueron censurados; los casos censurados se refieren a los estudiantes que se retiraron del estudio debido a diferentes razones.

### **5.1.2. Resultado de la estimación de la función de supervivencia mediante el método de Kaplan-Meier**

En esta sección se presentan los resultados obtenidos mediante la aplicación del método de Kaplan-Meier para estimar la función de supervivencia de los estudiantes matriculados en la E.P Matemática de la UNSAAC durante el periodo 2015-2023. El estudio incluyó un total de 326 estudiantes. El tiempo de supervivencia se define como el tiempo desde la matricula del alumno hasta el evento de interés (desistimiento) o hasta el final del periodo de estudio, en caso de no haber desertado.

En este contexto, la censura se refiere a aquellos estudiantes que no desertaron durante el periodo de observación, ya sea porque completaron sus estudios o porque aún están matriculados al finalizar el estudio. Estos estudiantes no experimentaron el evento de interés (desistimiento), pero su tiempo de seguimiento se considera hasta el último momento en el que se les tiene registro.

A continuación, se presentan las tablas y figuras de los resultados obtenidos para estimar la función de supervivencia, los cuales se realizaron haciendo uso del software R y se utilizó un nivel de confianza del 95% y un nivel de significancia del 5%

#### **Tabla 8**

*Los tiempos de supervivencia según el método de Kaplan-Meier por semestre académico de los estudiantes matriculados durante el periodo 2015-2023 de la E.P Matemática de la UNSAAC*

Tiempo	Nº Riesgo	Nº Evento	Supervivencia	Std.err	L.I 99%	L.S 99%
0	326	31	0.90491	0.01625	0.873619	0.9373
1	267	39	0.77273	0.02398	0.727132	0.8212
2	228	3	0.76256	0.02437	0.71626	0.8119
3	225	20	0.69478	0.0265	0.644729	0.7487
4	205	38	0.56599	0.02867	0.512507	0.6251
5	167	2	0.55921	0.02872	0.505663	0.6184

6	165	24	0.47787	0.02895	0.424377	0.5381
7	141	6	0.45754	0.02888	0.404294	0.5178
8	135	3	0.44737	0.02883	0.394288	0.5076
9	132	6	0.42704	0.02869	0.37435	0.4871
10	126	43	0.2813	0.02613	0.234487	0.3375
11	83	3	0.27113	0.02583	0.22495	0.3268
12	80	4	0.25758	0.02541	0.212286	0.3125
13	76	20	0.18979	0.0228	0.149974	0.2402
14	56	1	0.1864	0.02265	0.146908	0.2365
15	55	9	0.1559	0.0211	0.119578	0.2033
16	46	15	0.10506	0.01784	0.07532	0.1466
17	31	7	0.08134	0.01591	0.055443	0.1193
18	24	7	0.05762	0.01356	0.036324	0.0914
19	17	6	0.03728	0.01103	0.020879	0.0666
20	11	1	0.03389	0.01053	0.018432	0.0623
21	10	5	0.01695	0.00751	0.007107	0.0404
22	5	1	0.01356	0.00673	0.005122	0.0359
23	4	1	0.01017	0.00584	0.003298	0.0313
24	3	2	0.00339	0.00338	0.000479	0.024
25	1	1	0	0	0	0

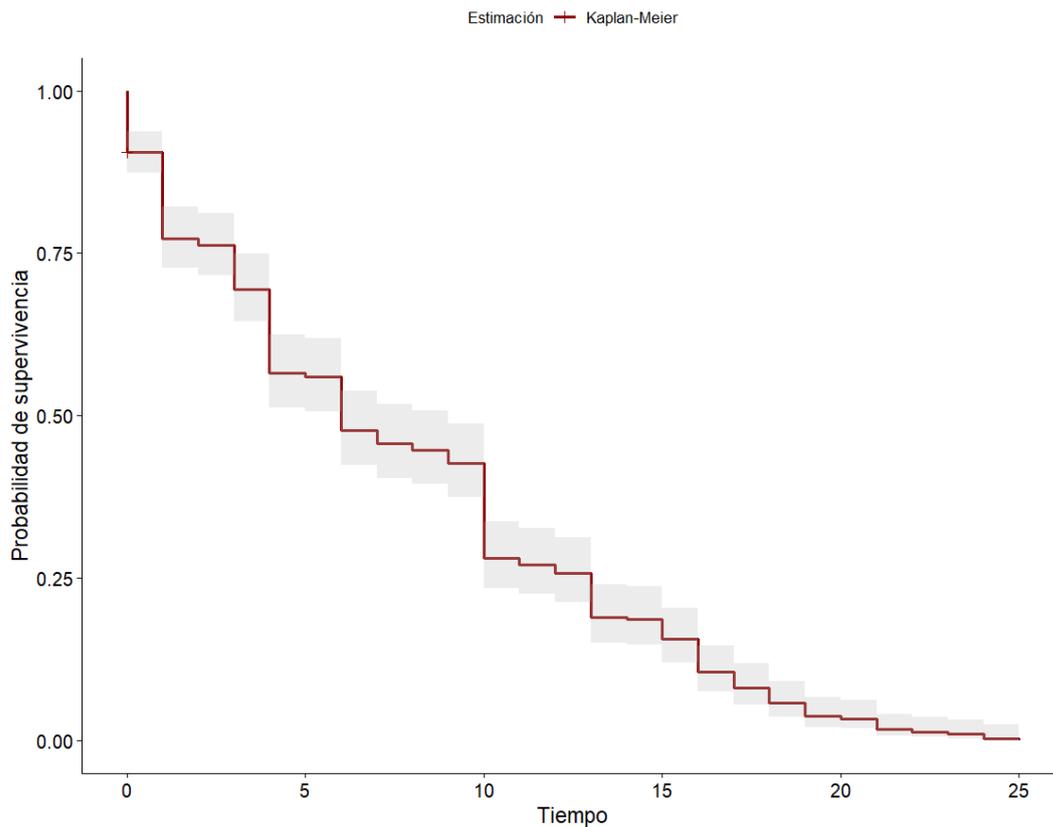
La Tabla 8 muestra los resultados del análisis de supervivencia de los estudiantes de la Escuela Profesional de Matemática de la UNSAAC (2015-2023) usando el método Kaplan-Meier. Se evalúa la probabilidad de que los estudiantes sigan matriculados hasta desertar o terminar la carrera.

Al inicio, 326 estudiantes estaban matriculados, con una supervivencia del 90.49% tras el primer semestre. Sin embargo, la retención disminuye progresivamente: al tercer semestre, la supervivencia cae al 69.48%, y al quinto, al 55.92%. Para el décimo semestre, solo el 28.13% de los estudiantes sigue matriculado, y al decimosexto, queda

solo el 10.51%. En el semestre 25, la supervivencia llega al 0%, indicando que todos han desertado o completado sus estudios.

#### Figura 4

*Curva de supervivencia de Kaplan-Meier- los estudiantes matriculados durante el periodo 2015-2023 de la E.P Matemática de la UNSAAC*



En la Figura 4 se puede apreciar la curva de supervivencia de los estudiantes matriculados durante el periodo 2015-2023 de la E.P Matemática de la UNSAAC. Esta gráfica nos permite estimar la probabilidad de que un alumno permanezca durante un período específico, en este caso, medido en periodos de semestre.

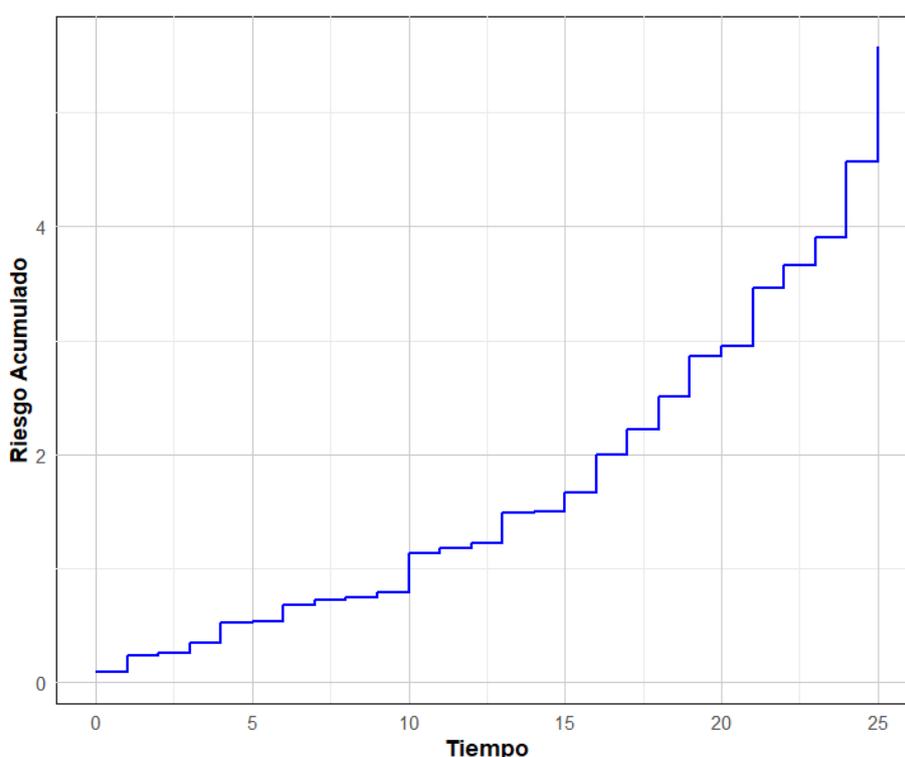
Las censuras están representadas por las marcas verticales a lo largo de la curva. Estas marcas indican los momentos en los que los datos de algunos estudiantes dejaron de ser observados debido a razones distintas al evento de interés (desistimiento).

La probabilidad de continuar los estudios es mayor durante los primeros semestres tras comenzar la universidad. Esto se debe a que los estudiantes matriculados pueden enfrentar diversas circunstancias que podrían comprometer su permanencia académica.

### 5.1.3. Estimación de la función de riesgo acumulado mediante el método de Nelson-Aalen

**Figura 5**

*Curva del riesgo acumulado según Nelson-Aalen - estudiantes matriculados durante el periodo 2015-2023 de la E.P Matemática de la UNSAAC*



La Figura 5 muestra que el riesgo acumulado de desistimiento crece de forma gradual durante los primeros cuatro semestres, para luego estabilizarse. Este comportamiento indica que los primeros semestres son cruciales para la retención de los estudiantes. Además, se observa un leve aumento en el riesgo acumulado durante el octavo semestre, posiblemente vinculado a la mayor exigencia académica de los cursos avanzados. Estos resultados subrayan la importancia de establecer programas de

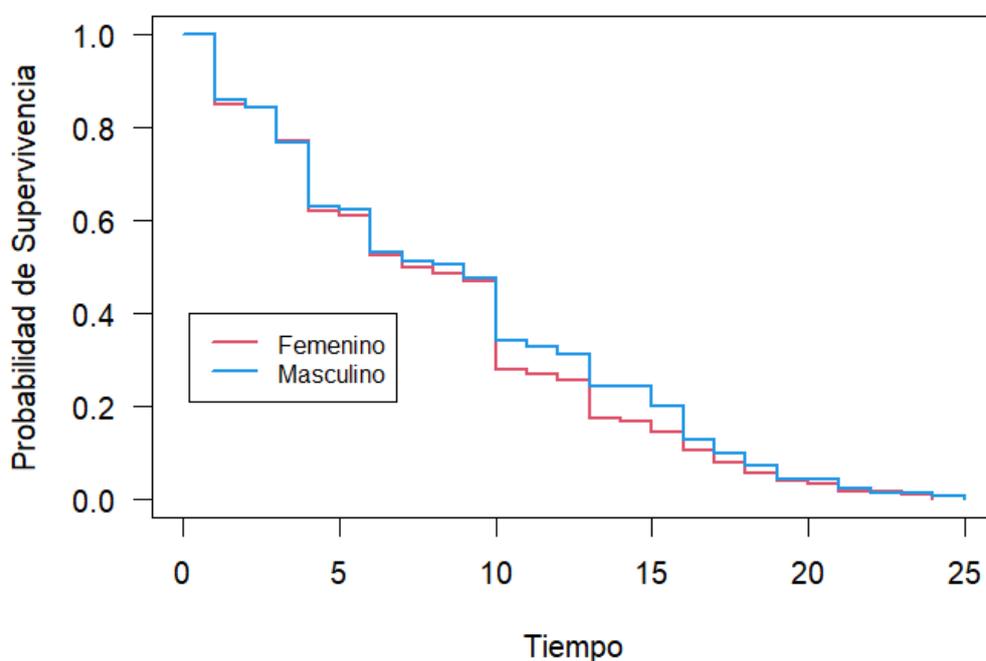
acompañamiento académico y orientación vocacional, especialmente al inicio de la trayectoria universitaria, con el objetivo de mejorar la retención estudiantil.

**5.1.4. Resultados de la comparación de las funciones de supervivencia según factores**

**5.1.4.1. Resultado de la prueba de Log-Rank para el sexo de los estudiantes matriculados durante el periodo de estudio de la E.P Matemática de la UNSAAC**

**Figura 6**

*Curva de supervivencia de los estudiantes matriculados durante el periodo 2015-2023 de la E.P Matemática de la UNSAAC según sexo*



La Figura 6 muestra las curvas de supervivencia de Kaplan-Meier para los estudiantes matriculados en la E.P. de Matemática de la UNSAAC durante el periodo 2015-2023, diferenciadas por sexo. Las curvas de los estudiantes femeninos y masculinos son similares, lo que indica que no hay diferencias significativas en la permanencia entre ambos grupos a lo largo del tiempo. A medida que transcurre el tiempo, la probabilidad de permanencia en el programa disminuye progresivamente para todos los estudiantes,

alcanzando casi cero al final del período, lo que refleja la finalización o abandono de los estudios.

### Hipótesis estadísticas

**Ho:** No hay diferencia significativa en la supervivencia entre los estudiantes matriculados en la EPM del sexo femenino y masculino.

**Ha:** Existe una diferencia significativa en la supervivencia entre los estudiantes matriculados en la EPM del sexo femenino y masculino.

### Tabla 9

*Resultados de la prueba de Log-Rank para el sexo de los estudiantes matriculados en la EPM durante 2015-2023.*

Resumen - prueba de Log-Rank -sexo de estudiantes						
	n	Observaciones	Esperados	Chi-Cuadrado	P_valor	Log-Rank
Femenino	159	147	135	1.064	0.302	2.4
Masculino	167	151	163	0.881	0.348	2.4

La Tabla 9 muestra los resultados de la prueba de Log-Rank para el sexo de los estudiantes matriculados en la EPM durante 2015-2023 de la UNSAAC.

Los resultados de la prueba de Log-Rank, presentados en la Tabla 8, indican que no hay diferencias significativas en las curvas de supervivencia entre estudiantes femeninos y masculinos matriculados en la EPM durante el periodo 2015-2023. Los p-valores obtenidos fueron de 0.302 para mujeres y 0.348 para hombres, ambos mayores que el nivel de significancia de 0.05. Esto sugiere que no se puede rechazar la hipótesis nula, lo que implica que el sexo de los estudiantes no tiene un impacto significativo en el evento estudiado (desistimiento) durante este periodo. Por tanto, se concluye que no

hay evidencia suficiente para afirmar que el sexo afecta las tasas de supervivencia de los estudiantes en este análisis.

**Tabla 10**

*Comparaciones globales*

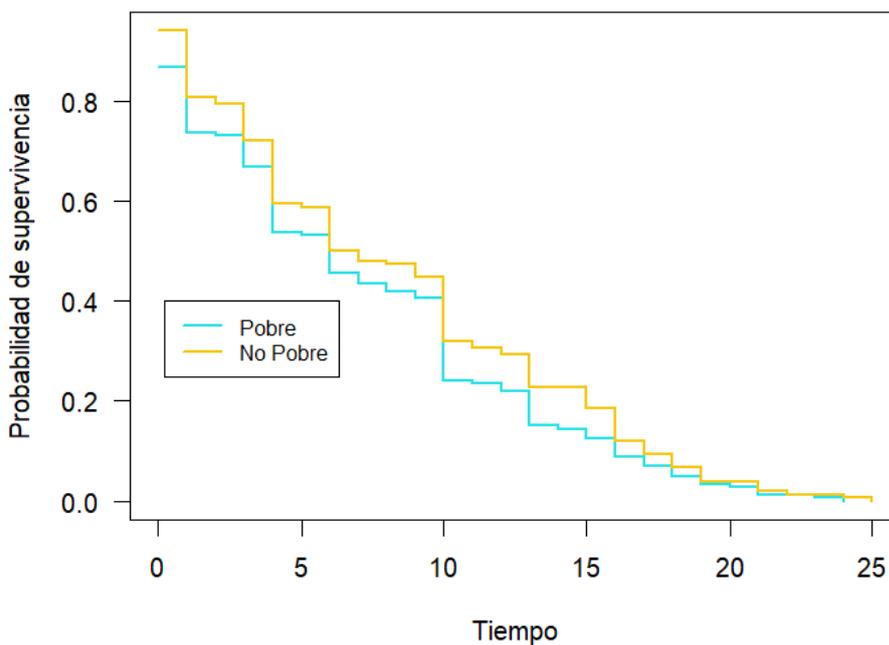
	Chi-Cuadrado	gl	P-value
Log-Rank	2.4	1	0.1

Con un nivel de significancia del 0.05 el p-value (0.1) > 0.05, por lo cual no se rechaza la hipótesis nula. Entonces, existe suficiente evidencia estadística para afirmar que no hay diferencia significativa en la supervivencia entre los estudiantes matriculados en la EPM durante 2015-2023 del sexo femenino y masculino.

### 5.1.4.2. Estadísticos de condición económica de los estudiantes matriculados durante el periodo 2015-2023 de la E.P Matemática de la UNSAAC

**Figura 7**

*Curva de supervivencia de los estudiantes matriculados durante el periodo 2015-2023 de la E.P Matemática de la UNSAAC según condición económica*



En la Figura 7 observamos, las curvas de supervivencia que indican que la condición económica (pobre o no pobre) no tiene un impacto significativo en la probabilidad de que los estudiantes permanezcan matriculados a lo largo del tiempo.

Aunque las curvas muestran una ligera variación, ambas siguen un patrón similar, indicando que otros factores podrían tener más peso en la permanencia de los estudiantes en la universidad.

#### **Hipótesis estadísticas**

**H<sub>0</sub>:** No hay diferencias significativas en las curvas de supervivencia de los estudiantes entre las dos categorías de condición económica (pobre y no pobre).

**Ha:** Existe una diferencia significativa en las curvas de supervivencia de los estudiantes entre las dos categorías de condición económica (pobre y no pobre).

**Tabla 11**

*Resultados de la prueba de Log-Rank para la condición económica de los estudiantes matriculados durante el periodo 2015-2023 de la E.P Matemática de la UNSAAC*

Resumen - Prueba de Log-Rank -Condición Económica						
	n	Observaciones	Esperados	Chi-Cuadrado	P-Valor	Log-Rank
Pobre	162	152	148	0.0915	0.7622	0.225
No Pobre	164	146	150	0.0907	0.7632	0.225

Según los resultados de la Tabla 11, se observa que tanto los estudiantes de condición económica pobre y no pobre tienen un p-valor mayor al nivel de significancia (0.05), lo que indica que no se rechaza la hipótesis nula ( $H_0$ ). Por lo tanto, existe suficiente evidencia estadística para afirmar que no hay diferencias significativas en las curvas de supervivencia de los estudiantes entre las dos categorías de condición económica (pobre y no pobre).

**Tabla 12**

*Comparaciones globales*

	Chi-Cuadrado	gl	P-value
Log-Rank	0.2	1	0.6

Según se muestra en la Tabla 12, la prueba estadística Log – Rank para la igualdad de las funciones de supervivencia, cuyo objetivo es comparar la función de supervivencia de la variable condición económica (0 =Pobre y 1 =No Pobre).

Considerando un nivel de significancia del 0.05 el p-value ( $0.6 > 0.05$ ), por lo cual no se rechaza la hipótesis nula. Entonces, existe suficiente evidencia estadística para afirmar que no hay diferencia significativa en la supervivencia entre los estudiantes de las dos categorías de condición económica (pobre y no pobre).

#### ***5.1.5. Resultado del modelo de regresión de Cox***

Los resultados del modelo de regresión de Cox presentados a continuación representan un avance significativo en la comprensión de los factores que influyen en el desistimiento de los estudiantes matriculados en la E.P de Matemática en la UNSAAC durante el periodo 2015-2023. El objetivo de este estudio es modelar la supervivencia utilizando -variables importantes como la edad y la cantidad de hermanos, las cuales se consideraron como variables numéricas. Además, se incluyeron variables categóricas como el sexo, la condición económica y el lugar de procedencia, clasificadas en diferentes categorías. Estas variables se seleccionaron cuidadosamente para determinar su impacto tanto individual como conjunto en el tiempo de supervivencia de los estudiantes. El proceso de construcción de los modelos de riesgos proporcionales de Cox en el presente estudio fue realizado de manera sistemática y estructurada, con el objetivo de analizar los factores asociados al desistimiento estudiantil. Se inició con un modelo básico que incluye el tiempo el estado y la edad, el siguiente modelo incluyó el tiempo, estado y el número de hermanos, posteriormente, se fueron incorporando variables adicionales, así como iteraciones, hasta completar un total de 57 modelos.

En una segunda etapa, se desarrollaron modelos más complejos con múltiples variables explicativas, identificando combinaciones significativas. Posteriormente, se analizaron interacciones entre variables (como Edad-CondEcon y Sexo-LugProced) para detectar efectos diferenciados en subgrupos.

Se construyeron 57 modelos iterativos, para garantizar parsimonia y robustez, evaluando significancia estadística.

Finalmente, se realizó la selección de modelos mediante los criterios de AIC y BIC. El modelo 33 fue identificado como el mejor modelo según ambos criterios, presentando los valores más bajos de AIC y BIC

**Tabla 13**

*Modelos de regresión de Cox*

<b>Modelos</b>	<b>AIC</b>	<b>BIC</b>
MdlCox1	2467.817	2474.11
MdlCox2	2812.354	2816.051
MdlCox3	2812.462	2816.159
MdlCox4	2814.661	2818.358
MdlCox5	2836.094	2891.551
MdlCox6	2471.791	2482.882
MdlCox7	2468.276	2479.367
MdlCox8	2470.622	2481.713
MdlCox9	2499.23	2613.84
MdlCox10	2814.392	2825.483
MdlCox11	2815.998	2827.09
MdlCox12	2853.077	2967.686
MdlCox13	2816.192	2827.283
MdlCox14	2844.776	2959.386
MdlCox15	2838.749	2953.359
MdlCox16	2474.083	2499.963
MdlCox17	2476.037	2501.917
MdlCox18	2530.056	2762.973
MdlCox19	2470.299	2496.179
MdlCox20	2516.354	2749.271
MdlCox21	2528.795	2761.712
MdlCox22	2821.533	2847.413
MdlCox23	2875.757	3108.674
MdlCox24	2871.474	3104.391
MdlCox25	2854.181	3087.098
MdlCox26	2480.424	2535.88
MdlCox27	2559.855	3021.991
MdlCox28	2748.626	3218.157
MdlCox29	2535.311	2978.963
MdlCox30	2893.18	3336.831
MdlCox31	2695.322	3638.081

MdlCox32	2469.791	2477.185
MdlCox33	2466.716	2471.515
MdlCox34	2468.998	2476.392
MdlCox35	2485.595	2544.749
MdlCox36	2812.483	2819.877
MdlCox37	2814.278	2821.672
MdlCox38	2835.466	2894.619
MdlCox39	2814.303	2821.697
MdlCox40	2836.364	2895.517
MdlCox41	2837.757	2896.911
MdlCox42	2468.713	2479.804
MdlCox43	2470.968	2482.059
MdlCox44	2487.526	2550.377
MdlCox45	2468.207	2479.298
MdlCox46	2483.586	2546.437
MdlCox47	2487.084	2549.935
MdlCox48	2814.415	2825.506
MdlCox49	2836.288	2899.139
MdlCox50	2837.261	2900.112
MdlCox51	2838.001	2900.852
MdlCox52	2470.203	2484.991
MdlCox53	2485.577	2552.125
MdlCox54	2489.015	2555.563
MdlCox55	2485.355	2551.902
MdlCox56	2838.046	2904.594
MdlCox57	2487.346	2557.591

En la Tabla 13 se observan los valores de AIC y BIC para los 57 modelos de Cox.

A continuación, se explican los detalles de la selección:

Para la selección de modelos basados en el AIC y BIC, el modelo 33 mostro un valor de 2466.716 y 2471.515 respectivamente. Dado que estos modelos ofrecieron valores similares bajo el criterio del AIC y BIC, se procedió a evaluar con base en estos valores, se concluyó que el modelo que mejor se adecua a nuestros datos es el modelo 33 que incluye las variables: Edad y Sexo.

Los cálculos para estimar la relación que hay entre un conjunto de variables explicativas, fijas de estudio en la probabilidad de supervivencia de desistimiento de estudiantes matriculados en la E.P de Matemática, fueron realizados en el software R.

**Tabla 14***Coefficientes de regresión de Cox para las variables del modelo*

Variables	$\beta$	$Exp(\beta)$	$se(\beta)$	z	p	Intervalo de confianza	
						Inferior	Superior
Edad	-0.3693	0.6912	0.0234	-15.779	< 2e-16	0.6602	0.7237
SexoMasculino	-0.2056	0.8142	0.1166	-1.763	0.0779	0.6478	1.0232

*Nota.*  $\beta$  es el coeficiente de regresión,  $Exp(\beta)$  es el riesgo relativo,  $se(\beta)$  es el error estándar del coeficiente de regresión,  $Z$  es el estadístico  $Z$  y  $p$  es la probabilidad de obtener resultados como los observados si la hipótesis nula fuera cierta.

En la Tabla 14 se observa los coeficientes de regresión de Cox para las variables del mejor modelo seleccionado, se empleó un enfoque riguroso para la selección del modelo, utilizando criterios estadísticos como AIC y BIC para identificar el modelo más adecuado. El estudio utilizó un modelo de regresión de Cox para analizar el riesgo de desistimiento en 326 estudiantes matriculados durante el periodo 2015-2023 de la E.P Matemática de la UNSAAC. Los estudiantes fueron seguidos desde su ingreso hasta el fin de su carrera, el evento de interés (desistimiento).

Los coeficientes revelan la relación entre las variables independientes (Edad y SexoMasculino) y el desistimiento estudiantil. El coeficiente de Edad (-0.3693) indica que, a mayor edad, menor es el riesgo de desistimiento, mientras que el coeficiente de SexoMasculino (-0.2056) indica que los hombres tienen una tasa de desistimiento ligeramente inferior a la de las mujeres, aunque esta diferencia no es muy significativa.

El riesgo relativo  $\text{Exp}(\beta)$  destaca que por cada año adicional de edad, el riesgo de desistimiento disminuye en un 30.88% ( $\text{Exp}(\beta) = 0.6912$ ). Para los hombres, el riesgo es un 18.58% menor que para las mujeres ( $\text{Exp}(\beta) = 0.8142$ ), aunque esta diferencia no es tan marcada.

La precisión de las estimaciones se refleja en los errores estándar: para Edad (0.0234) es alta, lo que respalda su efecto claro sobre el desistimiento, mientras que para SexoMasculino (0.1166) la precisión es menor. El valor z de Edad (-15.779) confirma su alta significancia, mientras que el de SexoMasculino (-1.763) indica que no es estadísticamente significativo.

El valor p de la variable Edad ( $p < 2e-16$ ) confirma su fuerte influencia en el desistimiento, en tanto que el de SexoMasculino ( $p = 0.0779$ ) no alcanza el umbral de significancia del 0.05. Los intervalos de confianza para Edad (0.6602, 0.7237) refuerzan la conclusión de que la edad reduce el riesgo de desistimiento, pero para SexoMasculino (0.6478, 1.0232) el impacto no es concluyente con un 95% de confianza.

De acuerdo al modelo de Cox definido previamente, el modelo seleccionado para analizar desistimiento de los estudiantes matriculados en la E.P de Matemática en la UNSAAC durante el periodo 2015-2023, es el siguiente:

$$h_i(t, X) = h_0(t)e^{-0.3693X_1 - 0.2056X_2}$$

Donde:

$X_1$  es la Edad.

$X_2$  es el SexoMasculino.

Además,  $h_0(t)$  es la función de riesgo base que es común para todos los individuos y no depende de las variables explicativas.

**Tabla 15***Índice de concordancia del modelo de riesgos proporcionales de Cox*

	Valor	se
Índice de concordancia	0.871	0.01

La Tabla 15 muestra el índice de concordancia, el cual mide la capacidad del modelo para predecir correctamente qué individuos experimentarán el evento (desistimiento) y cuáles no. Un valor de 0.871 indica que el modelo tiene una sólida capacidad de discriminación, lo que significa que en el 87.1 % de los casos, predice con precisión que un estudiante con mayor riesgo de desistimiento abandonará antes que uno con menor riesgo.

**Tabla 16***Contraste de hipótesis del modelo de Cox*

	Chi-Cuadrado	gl	P-value
Prueba de razón de verosimilitud	350.1	2	<2e-16
Test de Wald	249.6	2	<2e-16
Prueba de puntuación (Log-Rank)	297.3	2	<2e-16

La Tabla 16 muestra los resultados del contraste de hipótesis del modelo de riesgos proporcionales de Cox aplicado para analizar el desistimiento de estudiantes de la E.P. de Matemática de la UNSAAC entre 2015 y 2023. Las tres pruebas presentadas confirman la validez del modelo. La prueba de razón de verosimilitud, con un estadístico chi-cuadrado de 350.1 y un valor p menor que  $2e-16$ , rechaza la hipótesis nula, lo que indica un buen ajuste del modelo y la relevancia de las variables. El test de Wald, con un chi-cuadrado de 249.6 y un valor p igualmente bajo, refuerza la significancia de los coeficientes y el aporte de las variables independientes a la predicción del desistimiento.

Finalmente, la prueba de puntuación (Log-Rank), con un chi-cuadrado de 297.3 y un valor  $p$  también menor que  $2e-16$ , sugiere diferencias significativas en las tasas de desistimiento según las variables, confirmando la adecuación del modelo para el período analizado.

## 5.2. Discusiones

Los resultados de este estudio sobre el desistimiento estudiantil en la Escuela Profesional de Matemática de la UNSAAC brindan una visión integral de los factores que inciden en este fenómeno, confirmando la necesidad de abordar el abandono en la educación superior. Coincidiendo con el estudio de Paz (2022) en Ingeniería Civil, se encontró que una gran parte de los abandonos ocurre durante el primer año, siendo los primeros cursos clave para la permanencia. El 58% de los desertores en su estudio no aprobó ninguna materia, lo que refuerza la importancia de ofrecer apoyo académico temprano.

Vélez (2020) resalta que los factores sociodemográficos y académicos influyen considerablemente en el desistimiento, lo cual también se observó en esta investigación, donde el rendimiento académico y las características personales fueron determinantes. El análisis de supervivencia, similar al empleado por Jiménez (2021), permitió identificar factores de riesgo como los económicos y de salud, coincidiendo con la realidad de los estudiantes.

El estudio de Paucar y Quispe (2021) confirma una tasa de desistimiento del 61.8% en la Escuela de Matemática, destacando la influencia de variables como el sexo, la edad y los ingresos familiares. Estos resultados reafirman que el desistimiento es un problema multidimensional que involucra no solo el rendimiento académico, sino también el contexto socioeconómico.

Basurto (2021) también destaca la influencia de factores socioeconómicos en el desistimiento, subrayando la necesidad de intervenciones que consideren la realidad económica de los estudiantes. Además, la propuesta de Medina y Romero (2020) para desarrollar un modelo de retención coincide con nuestras recomendaciones de implementar programas de apoyo académico y psicológico.

Vélez (2020) también menciona la importancia de alertas tempranas para identificar a los estudiantes en riesgo y facilitar estrategias de retención más efectivas. Aunque este estudio ofrece una perspectiva valiosa, sus resultados están limitados a una sola institución, por lo que futuros estudios podrían incluir comparaciones con otras universidades para obtener una visión más amplia sobre el desistimiento en la educación matemática.

Los hallazgos subrayan la necesidad urgente de abordar el desistimiento estudiantil desde una perspectiva integral, considerando tanto los factores académicos como los socioeconómicos. La implementación de estrategias de retención basadas en estos resultados permitirá no solo reducir la tasa de abandono, sino también mejorar la calidad educativa y el bienestar de los estudiantes en la UNSAAC.

## CONCLUSIONES

El análisis utilizando el modelo de riesgos proporcionales de Cox demostró que factores como el sexo y la edad son variables significativas que influyen en el desistimiento de los estudiantes de la Escuela Profesional de Matemática de la UNSAAC. Esto indica que mejorar estas condiciones puede ser crucial para aumentar la retención estudiantil.

La estimación de la función de supervivencia mediante el método de Kaplan-Meier mostró que la mayoría de los estudiantes que desisten lo hacen durante los primeros años de su carrera. Esto resalta la importancia de identificar a los alumnos en riesgo desde el inicio y ofrecer apoyo adicional para aumentar las tasas de retención.

El análisis de la función de riesgo acumulado evidenció que el riesgo de desistimiento se acumula a lo largo del tiempo, con un aumento significativo en las tasas de desistimiento en los primeros semestres. Este hallazgo indica la necesidad de implementar estrategias de intervención tempranas que aborden las causas del desistimiento desde el principio de la formación académica.

La aplicación de la prueba de log-rank permitió identificar diferencias significativas en las curvas de desistimiento entre diferentes grupos de estudiantes, según variables demográficas y académicas. Estos resultados sugieren que las intervenciones para mejorar la retención deben ser personalizadas, teniendo en cuenta las características específicas de cada grupo en riesgo de desistimiento.

## RECOMENDACIONES

Se sugiere realizar investigaciones adicionales que profundicen en los factores que contribuyen al desistimiento de estudiantes en la Escuela Profesional de Matemática, incorporando un conjunto más amplio de variables. Esta ampliación debe incluir aspectos como el clima institucional, el nivel de satisfacción académica, la calidad de la enseñanza y las dinámicas de grupo entre los estudiantes. La integración de estas variables proporcionará una comprensión más integral de los factores que influyen en el desistimiento, permitiendo el diseño de intervenciones más efectivas y ajustadas a las necesidades de los alumnos.

A la dirección de la Escuela Profesional de Matemática se le recomienda implementar un sistema de apoyo académico estructurado que incluya tutorías personalizadas, talleres de habilidades de estudio y programas de mentoría. Estas medidas facilitarán la adaptación de los estudiantes a las exigencias académicas y promoverán su permanencia en la carrera.

Se recomienda fomentar un ambiente académico inclusivo y colaborativo, donde los estudiantes se sientan valorados y apoyados. Esto implica promover actividades extracurriculares que fortalezcan el sentido de pertenencia y la interacción social, como grupos de estudio, conferencias y seminarios, que no solo enriquezcan su formación académica, sino que también fortalezcan la comunidad estudiantil.

Además, se sugiere a las autoridades académicas de la UNSAAC evaluar periódicamente la efectividad de las estrategias implementadas para abordar el desistimiento en la E.P de matemática. Esto permitirá realizar ajustes necesarios basados en la retroalimentación de los estudiantes y los resultados obtenidos, asegurando que las intervenciones sean pertinentes y efectivas para mejorar la retención estudiantil.

Por último, se recomienda establecer políticas institucionales que aborden las barreras económicas que enfrentan algunos estudiantes, tales como becas, subsidios y apoyo financiero. Estas iniciativas ayudarán a garantizar que todos los estudiantes, independientemente de su situación económica, tengan la oportunidad de completar sus estudios en la Escuela Profesional de Matemática

## BIBLIOGRAFÍA

- Allison, P. (2014). *Event history and survival analysis: Regression for longitudinal event data*. SAGE Publications.
- Alvarado, E., Morales, D., y Ortiz, J. (2020). Determinantes de la probabilidad de robo a micronegocios en el área metropolitana de Monterrey. *Revista de Economía*, 37(94). <https://doi.org/https://doi.org/10.33937/reveco.2020.129>
- Álvarez, M., Álvarez, C., y Gómez, A. (2022). Evaluación de supervivencia en el adulto mayor con diabetes mellitus de acuerdo con la Guía de Práctica Clínica. *Atención familiar*, 29(3), 154-159.  
<https://doi.org/http://dx.doi.org/10.22201/fm.14058871p.2022.3.82833>
- Álvarez, M., Álvarez, C., y Gómez, A. (2022). Evaluación de supervivencia en el adulto mayor con diabetes mellitus de acuerdo con la Guía de Práctica Clínica. *Atención Familiar*, 29(3).  
<https://doi.org/https://doi.org/10.22201/fm.14058871p.2022.3.82833>
- Alvarez, R. (2023). *Análisis de la Supervivencia: Regresión de Cox*. Amazon Digital Services.
- Arteaga, J., & Moreno, J. (2020). Determinantes de la estructura de largo plazo del mercado de Afores en México: un análisis de supervivencia. *Entreciencias: diálogos en la sociedad del conocimiento*, 8(22).  
<https://doi.org/https://doi.org/10.22201/enesl.20078064e.2020.22.72640>
- Basurto, I. (2021). *Los factores socioeconómicos, académicos e individuales y su relación con la deserción estudiantil en una universidad privada de Lima- periodo 2019*. Lima: Universidad Privada Norbert Wiener.  
[https://repositorio.uwiener.edu.pe/bitstream/handle/20.500.13053/5731/T061\\_07345898\\_D.pdf?sequence=1&isAllowed=y](https://repositorio.uwiener.edu.pe/bitstream/handle/20.500.13053/5731/T061_07345898_D.pdf?sequence=1&isAllowed=y)

- Boj del Val, E. (2023). *El modelo de regresión de Cox*. Universidad de Barcelona.
- Campos, N. (2021). Supervivencia de pacientes con cáncer de próstata en un hospital de Bogotá, Colombia 2008-2014. *Duazary*, 18(3), 259-268.  
<https://doi.org/https://doi.org/10.21676/2389783X.4238>
- Cerron, L., & Saenz, R. (2023). *Factores de riesgo de la deserción escolar en tiempos de covid 19 en la I.E. N° 30069 Callaballauri-Chupaca*. Huancayo: Universidad nacional del centro del Perú.  
[https://repositorio.uncp.edu.pe/bitstream/handle/20.500.12894/9572/T010\\_70212159\\_T.pdf?sequence=1&isAllowed=y](https://repositorio.uncp.edu.pe/bitstream/handle/20.500.12894/9572/T010_70212159_T.pdf?sequence=1&isAllowed=y)
- Cocoletzi, A., Solano, F., Vázquez, V., y Velasco, F. (2020). Análisis del modelo de Cox con una base de datos de crédito. *XIII Semana Internacional de la Estadística y la Probabilidad*.  
<https://cape.fcfm.buap.mx/SIEP/2020/carteles/cartel23.pdf>
- Córdova, P., y Yurevna, M. (2019). Modelo de Cox multinivel estratificado con interacción por índice de riqueza para análisis de riesgo de muerte de niños menores de cinco años - ENDSA 2008. *Investigación y Desarrollo*, 19(2).  
<https://doi.org/10.23881/idupbo.019.2-7e>
- Felipe, L., Flores, M., y Cabezas, A. (2021). El análisis de supervivencia en la gestión del riesgo de crédito: El caso de una cartera hipotecaria ecuatoriana. *Perfiles*, 1(26). <https://doi.org/https://doi.org/10.47187/perf.v1i26.137>
- Gómez, G., y Cadarso, C. (2017). El modelo de riesgos proporcionales de Cox y sus extensiones. Impacto en Estadística y Biomedicina. *La Gaceta de la RSME*, 20(3), 513-538.
- Guillermo, A., Pesci, S., Wright, R., Bartel, E., Marín, L., Ignacio, J., . . . Ceriani, L. (2022). Sobrevida en pacientes internados en unidades de cuidados intensivos por

- covid-19 en la provincia de Buenos Aires, Argentina. *Revista Argentina de Salud Pública*, 14(1). [https://www.scielo.org.ar/scielo.php?pid=S1853-810X2022000200048&script=sci\\_arttext](https://www.scielo.org.ar/scielo.php?pid=S1853-810X2022000200048&script=sci_arttext)
- Harrell, F., Lee, K., y Mark, D. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, 15(4), 361-387.
- Hernández, R., & Mendoza, C. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. Mc Graw Hill Education.
- Hernández, Z. (2012). *Métodos de análisis de datos: apuntes*. Iberus. [https://www.unirioja.es/cu/zehernan/docencia/MAD\\_710/Lib489791.pdf](https://www.unirioja.es/cu/zehernan/docencia/MAD_710/Lib489791.pdf)
- Horbath, J. (2022). Análisis de supervivencia de pacientes indígenas mexicanos contagiados con COVID-19 iniciando la pandemia. *Revista Latinoamericana de Población*(16). <https://dialnet.unirioja.es/servlet/articulo?codigo=8224400>
- Hosmer, D., Lemeshow, S., y May, S. (2008). *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*. John Wiley & Sons.
- Hosmer, D., Lemeshow, S., y May, S. (2011). *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*. Wiley.
- Jiménez, A. (2022). *Deserción estudiantil en Educación Técnica Productiva del CETPRO N°005 “El Tablazo” Corrales, Tumbes, 2021*. Tumbes: Universidad Alas Peruanas. [https://repositorio.uap.edu.pe/bitstream/handle/20.500.12990/12794/Tesis\\_desrci%3%b3n\\_estudiantil\\_educaci%3%b3n\\_t%3%a9cnica\\_Cetpro%20N%2%b0005\\_El%20Tablazo\\_Corrales\\_Tumbes.pdf?sequence=1&isAllowed=y](https://repositorio.uap.edu.pe/bitstream/handle/20.500.12990/12794/Tesis_desrci%3%b3n_estudiantil_educaci%3%b3n_t%3%a9cnica_Cetpro%20N%2%b0005_El%20Tablazo_Corrales_Tumbes.pdf?sequence=1&isAllowed=y)
- Jiménez, L. (2021). Factores asociados a la deserción en la formación doctoral. El caso de la Universidad de los Andes, Colombia. *Universidad de los Andes*.

<https://repositorio.uniandes.edu.co/server/api/core/bitstreams/cd33f50e-d4b5-44a4-87d8-be3e72ee0989/content>

Klein, J., y Moeschberger, M. (2003). *Survival Analysis: Techniques for censored and truncated data*. Springer. <https://link.springer.com/book/10.1007/b97377>

Klein, J., y Moeschberger, M. (2013). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer.

Klein, J. P., Moeschberger, M. L., & Therneau, T. M. (2014). *Survival analysis: Techniques for censored and truncated data (2nd ed.)*. Springer.

Kleinbaum, D., y Klein, M. (2012). *Survival Analysis*. Springer.

Martín, A. (2022). Sobrevivamos al análisis de supervivencia (primera parte).

*Angiología*, 74(4). <https://doi.org/https://dx.doi.org/10.20960/angiologia.00393>

Martín, A. (2022). Sobrevivamos al análisis de supervivencia (segunda parte).

*Angiología*, 74(5). <https://doi.org/https://dx.doi.org/10.20960/angiologia.00394>

Medina, M., & Romero, T. (2020). Modelo de supervivencia para analizar el abandono escolar de la URACCAN-Nueva Guinea. *Revista Ciencia e Interculturalidad*, 26(1), 18-35.

Molina, M. (2015). Hazard ratio: cuando el riesgo varía a lo largo del tiempo. *Pediatría Atención Primaria*, 17(66), 185-188.

<https://doi.org/https://dx.doi.org/10.4321/S1139-76322015000300023>

Neis, E., Covinich, M., y Scipioni, G. (2021). Análisis sensorial de mate frío (tereré) de jugo artificial sabor naranja mediante estadística de supervivencia. *Revista de Ciencia y Tecnología*(35). [https://www.scielo.org.ar/scielo.php?pid=S1851-75872021000100011&script=sci\\_arttext](https://www.scielo.org.ar/scielo.php?pid=S1851-75872021000100011&script=sci_arttext)

Ortega et al. (2023). Estudios de supervivencia. Modelo de riesgos proporcionales. Regresión de Cox. *AEPap*, 19(4).

<https://evidenciasenpediatria.es/articulo/8249/estudios-de-supervivencia-modelo-de-riesgos-proporcionales-regresion-de-cox>

Ortega, E., Ochoa, C., y Molina, M. (2023). Estudios de supervivencia. Modelo de riesgos proporcionales. Regresión de Cox. *Evidencias en pediatría*, 19(48), 1-6.  
[https://evidenciasenpediatria.es/files/41-14457-RUTA/Fund\\_48\\_RegresionCox.pdf](https://evidenciasenpediatria.es/files/41-14457-RUTA/Fund_48_RegresionCox.pdf)

Paucar, J., & Quispe, M. (2021). *Análisis de supervivencia: deserción estudiantil en la Escuela Profesional de Matemática de la Universidad Nacional de San Antonio Abad del Cusco*. Cusco: Universidad de San Antonio Abad del Cusco.  
<https://repositorio.unsaac.edu.pe/handle/20.500.12918/5846>

Paz, H. (2022). Aplicación de técnicas de análisis de supervivencia al estudio de la deserción estudiantil en una carrera de ingeniería civil. *Scielo Preprints*, 1-26.  
<https://doi.org/https://doi.org/10.1590/SciELOPreprints.4589>

Perú. (2014). Ley N° 30220, Ley Universitaria. Diario Oficial El Peruano. Recuperado de <https://www.elperuano.pe>

Quiñones, L., Jara, D., Alvarado, N., Milla, M., & Gamarra, O. (2020). Modelo para la estimación de la deserción estudiantil Awajún y Wampis empleando minería de datos. *RECyT*, 22(34), 45-50.  
[https://rid.unam.edu.ar/bitstream/handle/20.500.12219/2839/Qui%c3%blones%20Huatangari%20L\\_2020\\_Modelo%20para%20la%20estimaci%c3%b3n.pdf?sequence=1&isAllowed=y](https://rid.unam.edu.ar/bitstream/handle/20.500.12219/2839/Qui%c3%blones%20Huatangari%20L_2020_Modelo%20para%20la%20estimaci%c3%b3n.pdf?sequence=1&isAllowed=y)

Ramírez, J., Regino, E., & Guerrero, S. (2017). Comparación de métodos de estimación en regresión de Cox. *Comunicaciones en Estadística*, 10(1), 101-112.

Ramírez, N., Castillo, J., Herrera, L., Gómez, B., Valbuena, A., y Acuña, L. (2021). Supervivencia en las personas que viven con VIH en el marco del sistema de

- salud colombiano 2011-2018. *Infectio*, 25(4).  
<https://doi.org/https://doi.org/10.22354/in.v25i4.960>
- Rossello, X., & González, M. (2022). Análisis de supervivencia en investigación cardiovascular (II): metodología estadística en situaciones complejas. *Revista Española de Cardiología*, 75(1), 77-85.  
<https://doi.org/https://doi.org/10.1016/j.recesp.2021.06.027>
- Rossello, X., y González, M. (2022). Análisis de supervivencia en investigación cardiovascular (I): lo esencial. *Revista Española de Cardiología*, 75(1), 67-76.  
<https://doi.org/https://doi.org/10.1016/j.recesp.2021.05.017>
- Rubio, E., y Martínez, T. (2016). *Manual de análisis de supervivencia: curvas de supervivencia y regresión de Cox*. Prensas de la Universidad de Zaragoza.
- Saavedra, E. (2023). *Aplicación de un modelo de riesgos competitivos bayesiano*. Lima: Pontificia Universidad Católica del Perú.  
[https://tesis.pucp.edu.pe/repositorio/bitstream/handle/20.500.12404/27235/SAAVEDRA\\_PALACIOS\\_ERICK\\_APLICACION\\_MODELO\\_RIESGOS.pdf?sequence=1&isAllowed=y](https://tesis.pucp.edu.pe/repositorio/bitstream/handle/20.500.12404/27235/SAAVEDRA_PALACIOS_ERICK_APLICACION_MODELO_RIESGOS.pdf?sequence=1&isAllowed=y)
- Sánchez, P. (2020). *Derecho procesal civil: Principios y prácticas*. Editorial Jurídica y Social. Recuperado de <https://www.editorialjuridica.com>.
- Salazar et al. (2020). *Introducció al anàlisi de supervivencia avanzada*. Bonaventuriana.  
[https://api.pageplace.de/preview/DT0400.9789588474939\\_A40559745/preview-9789588474939\\_A40559745.pdf](https://api.pageplace.de/preview/DT0400.9789588474939_A40559745/preview-9789588474939_A40559745.pdf)
- Salazar, J., García, E., Gaviria, C., y Guarín, V. (2020). *Introducció al anàlisi de supervivencia avanzada*. Bonaventuriana.

[https://api.pageplace.de/preview/DT0400.9789588474939\\_A40559745/preview-9789588474939\\_A40559745.pdf](https://api.pageplace.de/preview/DT0400.9789588474939_A40559745/preview-9789588474939_A40559745.pdf)

Sanchez, H., & Reyes, C. (2018). *Manual de términos en investigación científica, tecnológica y humanística*. <https://www.urp.edu.pe/pdf/id/13350/n/libro-manual-de-terminos-en-investigacion.pdf>

Steyerberg, E. (2019). *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Springer.

Torreglosa, S., Grisales, H., Morales, E., Hernández, J., Huerta, R., Barquet, S., y Palacio, L. (2022). Supervivencia y factores asociados en pacientes con cáncer cervicouterino atendidas por el Seguro Popular en México. *salud publica mexicana*, 64(1), 76-86. <https://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=104723>

UNESCO. (2015). Deserción estudiantil. Definición y tipos. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. Recuperado de <https://www.unesco.org>

Vega, J. (2019). El análisis de supervivencia como técnica para la evaluación de la validez predictiva en la psicología jurídica. *Anuario de Psicología jurídica*, 29(1), 1-10. <https://doi.org/https://doi.org/10.5093/apj2018a11>

Vélez, D. (2020). *Análisis de alertas tempranas sobre deserción estudiantil en educación superior : un estudio de caso para Medellín*. Medellín: Universidad Nacional de Colombia. <https://repositorio.unal.edu.co/handle/unal/86763>

Zuluaga, P., Useche, D., y Rojas, S. (2022). Relevancia, evolución y tendencias de la supervivencia empresarial. Una revisión de literatura en finanzas. *Tendencias*, 24(1). [http://www.scielo.org.co/scielo.php?pid=S0124-86932023000100252&script=sci\\_arttext](http://www.scielo.org.co/scielo.php?pid=S0124-86932023000100252&script=sci_arttext)

## **ANEXOS**

## A. Matriz de consistencia

Problemas	Objetivos	Hipótesis	Variables	Metodología
Problema General	Objetivo General	Hipótesis General	Variables Independientes	<b>Enfoque:</b> Cuantitativo <b>Tipo:</b> Aplicada <b>Diseño:</b> No experimental <b>Población:</b> Todos los estudiantes matriculados en la E.P de Matemática el periodo especificado, es decir, entre el 2015 y el 2023 (326 estudiantes matriculados) <b>Población de estudio:</b> 326 estudiantes matriculados en la E.P de Matemática el periodo especificado, es decir, entre el 2015 y el 2023 <b>Técnicas e Instrumentos para la Recolección de datos:</b> Observación documental de las listas de los estudiantes matriculados por semestre, proporcionada por el centro de cómputo de la UNSAAC, en un período comprendido entre los años 2015 -2023
¿Cómo se puede explicar el desistimiento estudiantil en la Escuela Profesional de Matemática de la UNSAAC durante el periodo 2015-2023 utilizando el modelo de riesgos proporcionales de Cox?	Explicar el desistimiento estudiantil en la Escuela Profesional de Matemática de la UNSAAC durante el periodo 2015-2023 utilizando el modelo de riesgos proporcionales de Cox.	El modelo de riesgos proporcionales de Cox explica de manera significativa el desistimiento estudiantil en la Escuela Profesional de Matemática de la UNSAAC durante el periodo 2015-2023.	<ul style="list-style-type: none"> <li>➤ Edad</li> <li>➤ Sexo</li> <li>➤ Condición económica</li> <li>➤ Lugar de Procedencia</li> <li>➤ Numero de hermanos</li> </ul>	
Problemas específicos	Objetivos específicos	Hipótesis Especificas	Variable Dependiente	
a) ¿Cuál es la probabilidad de que los estudiantes de la Escuela Profesional de Matemática continúen sus estudios a lo largo del tiempo durante el periodo 2015-2023?  b) ¿Cómo se acumula el riesgo de desistimiento a medida que transcurre el tiempo durante el periodo 2015-2023?  c) ¿Existen diferencias significativas en las curvas de supervivencia del desistimiento estudiantil según variables categóricas como el sexo, el lugar de procedencia o la condición socioeconómica?	a) Estimar la probabilidad de que los estudiantes continúen sus estudios a lo largo del tiempo durante el periodo 2015-2023.  b) Estimar la acumulación del riesgo de desistimiento estudiantil conforme transcurre el tiempo durante el periodo 2015-2023.  c) Comparar las curvas de supervivencia del desistimiento estudiantil según variables categóricas como el sexo, el lugar de procedencia y la condición socioeconómica.  d)	a) La probabilidad de que los estudiantes continúen sus estudios varía significativamente a lo largo del tiempo durante el periodo 2015-2023.  b) El riesgo de desistimiento estudiantil se acumula significativamente conforme transcurre el tiempo durante el periodo 2015-2023.  c) El sexo, el lugar de procedencia y la condición socioeconómica contribuyen significativamente al desistimiento estudiantil en la Escuela Profesional de Matemática de la Facultad de Ciencias Químicas, Físicas y Matemática de la UNSAAC durante el periodo 2015-2023.	Tiempo desde la primera matricula del estudiante hasta el desistimiento  Evento: Desistimiento de los estudiantes matriculados	

## B. Tiempo de desistimiento y los semestres académicos (2015-2023)

TIEMPO	semestres
1	2015-1
2	2015-2
3	2015-2V
4	2016-1
5	2016-2
6	2016-2V
7	2017-1
8	2017-2
9	2017-2V
10	2018-1
11	2018-2
12	2018-2V
13	2019-1
14	2019-2
15	2019-2V
16	2020-1
17	2020-2
18	2020-2V
19	2021-1
20	2021-2
21	2021-2V
22	2022-1
23	2022-2
24	2022-2V
25	2023-1
26	2023-2

La tabla establece una relación entre los periodos de tiempo y los semestres académicos desde 2015 hasta 2023, considerando que cada año académico se divide en tres semestres:

- **Semestre regular 1** (inicio del año)
- **Semestre regular 2** (mitad del año).
- **Semestre vacacional (V)** (tras el segundo semestre).

El tiempo se mide en unidades consecutivas de semestres, asignando el valor de **1** al semestre 2015-1 y continuando hasta el **26**, correspondiente al semestre 2023-2. En total, se contemplan 26 semestres distribuidos linealmente.

La columna de semestres detalla los nombres de cada periodo:

- **"2015-1"**: primer semestre regular de 2015.
- **"2015-2"**: segundo semestre regular de 2015.
- **"2015-2V"**: periodo vacacional posterior al segundo semestre de 2015.

Este patrón se repite uniformemente hasta el semestre final incluido, 2023-2.

### C. Data utilizada para la investigación

Tiempo	Estado	Edad	Sexo	CondEcon	NumHerm	LugProced
24	1	35	2	1	5	2
3	1	23	1	0	3	2
15	1	30	1	0	2	10
23	1	28	1	1	4	4
0	1	32	1	0	4	2
10	1	23	2	0	4	4
3	1	20	2	1	4	10
21	1	26	2	0	4	2
22	1	32	2	1	5	5
4	1	21	1	0	6	3
6	1	19	1	1	4	12
1	1	18	1	1	2	3
6	1	21	2	1	4	9
25	1	30	2	1	3	3
9	1	22	2	1	3	9
0	1	17	2	1	4	9
0	0	21	2	1	3	13
0	0	20	1	1	4	15
0	1	18	1	0	1	4
0	1	17	2	1	2	5
0	0	19	2	0	2	5
0	1	17	1	1	1	12
0	1	18	2	0	6	4
3	1	19	1	0	1	12
4	1	20	1	1	2	14
4	1	21	2	0	4	13
0	0	21	1	0	1	3
18	1	33	2	0	1	16
4	1	19	2	0	5	12
12	1	27	1	1	6	4
0	0	33	2	1	2	3
4	1	17	1	0	5	7
19	1	29	2	0	3	13
0	1	17	1	1	4	4
3	1	22	2	0	1	7
24	1	31	1	0	2	8
10	1	29	2	1	6	3
4	1	17	2	1	2	4
17	1	28	2	0	3	8
21	1	30	1	1	5	5
5	1	20	1	1	1	1
14	1	32	1	0	3	11
0	1	17	2	0	2	13
0	0	20	2	1	3	6

2	1	21	2	0	5	16
8	1	28	1	0	3	4
2	1	18	1	0	6	14
0	1	17	1	0	2	6
17	1	34	2	0	3	7
17	1	28	1	1	3	7
20	1	34	1	1	5	6
17	1	29	2	1	5	12
21	1	28	1	0	5	10
6	1	21	1	1	6	4
21	1	30	2	0	3	9
21	1	28	2	1	5	13
18	1	27	1	1	6	13
17	1	27	1	0	3	14
11	1	25	1	1	5	8
17	1	31	1	0	4	14
17	1	28	2	0	3	15
6	1	20	1	1	1	13
6	1	17	2	0	4	5
18	1	28	1	0	4	1
15	1	29	2	0	3	13
18	1	30	2	1	6	9
2	1	19	2	1	4	12
15	1	29	1	0	5	4
18	1	30	1	1	2	5
9	1	19	2	0	2	10
3	1	20	1	0	3	3
3	1	18	2	1	6	5
3	1	20	2	1	5	8
3	1	22	2	1	2	11
19	1	30	2	0	5	1
1	1	17	1	0	4	5
1	1	18	1	1	5	13
13	1	26	2	0	3	7
12	1	28	1	0	2	6
19	1	30	1	0	4	14
19	1	29	1	1	2	6
3	1	18	2	0	3	2
16	1	28	2	1	4	6
19	1	30	2	0	4	5
13	1	28	1	0	2	2
18	1	29	2	0	4	13
16	1	32	2	1	6	10
18	1	29	2	1	4	11
9	1	23	1	0	4	12
0	1	18	1	1	2	9
0	0	19	2	1	1	2
19	1	30	2	0	5	14

6	1	22	1	1	1	5
0	0	20	2	1	6	2
0	0	19	1	0	4	4
16	1	28	2	1	1	3
16	1	30	2	1	2	6
13	1	29	1	1	5	2
9	1	22	2	0	4	3
10	1	24	1	0	4	2
0	0	19	1	1	3	11
12	1	28	2	0	2	12
16	1	28	1	0	4	9
7	1	20	2	1	2	10
16	1	33	2	1	3	10
0	1	18	1	1	3	6
16	1	28	2	0	4	14
15	1	26	2	0	2	1
4	1	19	1	1	1	11
16	1	32	1	0	4	11
16	1	30	1	0	2	1
1	1	18	2	1	4	5
16	1	28	1	1	1	2
1	1	20	1	0	6	14
6	1	22	1	0	1	11
15	1	29	2	1	4	10
7	1	21	1	1	5	6
3	1	19	2	0	2	5
16	1	34	1	0	1	10
1	1	21	2	0	2	4
13	1	28	1	0	1	12
16	1	29	2	0	4	7
3	1	19	2	0	3	3
1	1	23	2	1	6	4
16	1	30	2	1	5	13
13	1	28	2	1	3	12
16	1	33	2	1	2	3
1	1	20	1	0	2	2
13	1	30	1	1	2	9
16	1	28	2	1	4	6
15	1	32	2	0	6	10
15	1	35	2	1	3	6
11	1	27	2	0	6	10
3	1	19	1	0	2	8
8	1	23	2	1	2	9
15	1	34	2	0	5	1
12	1	27	2	1	4	3
15	1	28	1	1	4	8
13	1	30	1	1	3	12
13	1	28	2	1	6	9

13	1	29	2	0	6	5
13	1	32	2	1	4	6
13	1	30	2	1	4	9
13	1	29	1	0	5	11
13	1	28	1	0	4	5
13	1	27	2	0	1	8
13	1	26	2	0	3	1
13	1	35	2	0	5	13
13	1	27	1	1	2	11
13	1	30	2	1	6	6
13	1	30	1	0	6	5
13	1	32	1	1	6	12
7	1	22	2	0	6	3
10	1	27	2	1	5	12
1	1	17	2	1	6	6
1	1	18	1	1	1	6
0	0	21	1	1	2	10
0	0	22	1	1	3	5
5	1	25	2	1	6	2
8	1	24	1	0	5	10
11	1	26	2	0	3	6
10	1	28	2	1	4	6
10	1	27	2	1	4	6
10	1	24	1	0	4	10
10	1	23	1	1	3	11
10	1	30	2	1	3	6
10	1	30	1	0	1	6
10	1	26	1	0	2	2
10	1	29	2	0	6	6
10	1	33	1	0	5	9
10	1	27	1	0	3	8
10	1	28	2	1	2	9
10	1	31	1	1	3	8
9	1	25	2	0	6	5
6	1	19	2	0	3	1
0	0	21	2	0	2	10
7	1	26	2	1	2	6
9	1	25	1	1	2	3
7	1	20	1	0	3	6
1	1	18	2	1	4	6
10	1	27	1	1	3	6
10	1	24	2	1	4	7
10	1	30	2	0	1	10
10	1	24	2	0	4	6
10	1	23	2	0	4	10
7	1	21	1	1	5	15
10	1	26	1	1	2	15
10	1	34	1	1	1	2

10	1	35	1	0	5	6
10	1	24	2	1	5	11
10	1	23	2	1	4	9
10	1	30	2	1	4	4
10	1	28	1	0	5	3
10	1	32	1	0	3	1
10	1	27	2	1	4	7
10	1	32	1	1	5	7
10	1	30	2	0	1	13
10	1	26	2	0	3	6
10	1	28	1	1	6	6
10	1	27	1	1	2	15
10	1	28	1	0	4	16
10	1	30	1	1	5	2
10	1	27	1	1	3	13
10	1	26	1	1	6	11
10	1	32	2	1	4	13
10	1	34	1	0	1	5
10	1	35	1	1	3	7
10	1	33	1	0	6	2
4	1	20	2	0	6	8
4	1	19	2	0	1	8
4	1	20	2	0	3	11
6	1	25	2	1	5	2
0	0	21	1	1	1	13
4	1	22	1	1	3	2
6	1	20	2	0	5	15
6	1	21	1	0	4	8
0	1	17	1	0	4	1
0	1	18	2	0	1	6
6	1	24	2	1	4	7
0	1	18	1	1	2	9
0	0	19	2	1	3	9
6	1	24	2	1	2	14
6	1	23	1	0	5	13
0	0	19	1	1	1	8
0	1	18	1	1	3	5
6	1	22	1	1	3	9
0	1	18	1	0	5	1
6	1	23	2	0	2	16
6	1	25	2	1	6	4
0	1	18	1	1	6	6
6	1	25	2	1	6	2
6	1	24	2	0	5	16
0	1	17	1	1	4	5
6	1	21	1	1	2	1
0	0	19	2	0	1	13
0	1	18	1	1	1	11

6	1	22	2	0	1	13
0	1	18	2	1	4	8
0	1	17	2	0	5	13
6	1	28	1	1	1	5
6	1	25	2	1	5	11
6	1	23	1	0	6	8
4	1	20	2	0	2	7
4	1	21	2	1	5	5
4	1	20	2	0	2	10
4	1	22	1	1	6	14
4	1	20	2	1	4	6
4	1	19	1	0	3	3
4	1	21	1	1	2	1
4	1	19	1	1	1	13
4	1	19	1	1	2	11
4	1	21	2	1	2	8
4	1	22	1	1	6	7
4	1	19	1	1	1	8
4	1	23	2	1	6	11
4	1	24	2	0	4	3
4	1	23	2	1	6	11
4	1	22	1	0	6	4
4	1	25	1	0	2	2
4	1	23	1	1	4	1
4	1	25	1	1	1	6
4	1	23	2	1	3	5
4	1	21	1	1	5	9
4	1	22	1	0	6	11
4	1	19	2	0	5	14
4	1	23	2	0	4	6
4	1	22	2	1	6	13
4	1	20	1	0	4	1
4	1	20	2	0	3	13
3	1	18	1	0	5	1
0	0	20	2	0	3	6
0	1	17	2	1	1	10
1	1	18	2	0	1	5
1	1	20	1	0	4	2
1	1	21	2	0	2	10
3	1	20	1	0	3	8
1	1	18	2	0	1	4
1	1	19	2	0	6	5
0	1	18	1	1	1	5
0	0	20	1	1	1	13
0	0	21	2	1	1	10
3	1	19	1	1	4	14
3	1	22	1	0	2	5
3	1	18	2	1	3	7

3	1	19	2	0	5	6
0	1	18	2	1	6	9
3	1	23	2	0	2	4
3	1	22	1	0	3	2
1	1	17	2	1	3	1
0	1	18	1	0	1	5
1	1	19	1	0	3	7
1	1	17	1	0	1	9
0	0	20	1	1	5	9
1	1	18	2	0	5	14
1	1	20	1	0	1	3
1	1	17	2	0	6	15
1	1	19	1	0	1	5
1	1	18	2	0	3	6
1	1	19	2	1	6	1
1	1	20	2	1	2	15
1	1	23	2	0	2	16
1	1	20	1	1	4	5
0	1	17	1	1	5	16
1	1	18	2	0	1	1
1	1	20	1	0	3	3
1	1	19	2	1	1	14
1	1	20	1	1	5	6
1	1	18	1	0	2	14
1	1	20	2	0	2	6
1	1	19	2	1	6	6
1	1	19	1	0	5	15
0	1	18	2	0	6	11
1	1	19	1	1	1	6
1	1	22	1	1	2	2
1	1	20	1	0	5	11
0	1	18	1	1	6	1
0	0	19	2	1	1	3
0	0	19	1	0	4	6
0	0	20	2	0	5	11
0	1	18	1	0	1	15
0	0	20	2	0	4	2
0	0	21	1	0	3	14
0	0	20	2	1	6	16
0	1	17	1	0	3	1
0	1	18	1	0	4	15
0	0	19	2	1	3	16

---

## GLOSARIO

E.P.M: Escuela Profesional de Matemática.

UNSAAC: Universidad Nacional San Antonio Abad Del Cusco.