

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



**TESIS**

---

**FACTORES QUE INFLUYEN EN LA DESERCIÓN UNIVERSITARIA DE LOS  
ESTUDIANTES DE LA ESCUELA PROFESIONAL DE MATEMÁTICA DE LA  
UNSAAC, CON EL MODELO DE REGRESIÓN LOGÍSTICA, CUSCO 2022-2023**

---

**PRESENTADA POR:**

Br. FRANKLIN CUCHILLO VALVERDE

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

**ASESORA:**

Dra. NELLY MARÍA SALAZAR PEÑA

**CUSCO – PERÚ**

**2024**

# INFORME DE ORIGINALIDAD

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

El que suscribe, Asesor del trabajo de investigación/tesis titulada: FACTORES QUE INFLUYEN EN LA DESERCIÓN UNIVERSITARIA DE LOS ESTUDIANTES DE LA ESCUELA PROFESIONAL DE MATEMÁTICA DE LA UNSAAC, CON EL MODELO DE REGRESIÓN LOGÍSTICA, CUSCO 2022-2023

presentado por: Bs. FRANKLIN CUCHILLO VALVERDE con DNI Nro.: 74401449 presentado por: ..... con DNI Nro.: ..... 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 DOS 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 5%.

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 página del reporte del Sistema Antiplagio.

Cusco, 22 de OCTUBRE de 2024

  
Firma

Post firma NELLY MARÍA SALAZOR PEÑA

Nro. de DNI 23910370

ORCID del Asesor 0000-0002-7582-9492

Se adjunta:

1. Reporte generado por el Sistema Antiplagio.
2. Enlace del Reporte Generado por el Sistema Antiplagio: oid: 27259:397396326

NOMBRE DEL TRABAJO

**FACTORES QUE INFLUYEN EN LA DESERCIÓN UNIVERSITARIA DE LOS ESTUDIANTES DE LA ESCUELA PROFESIONAL DE**

AUTOR

**FRANKLIN CUCHILLO VALVERDE**

RECUENTO DE PALABRAS

**18714 Words**

RECUENTO DE CARACTERES

**110776 Characters**

RECUENTO DE PÁGINAS

**95 Pages**

TAMAÑO DEL ARCHIVO

**3.6MB**

FECHA DE ENTREGA

**Oct 22, 2024 6:42 PM GMT-5**

FECHA DEL INFORME

**Oct 22, 2024 6:43 PM GMT-5****● 5% 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
- 2% Base de datos de trabajos entregados
- 0% Base de datos de publicaciones
- Base de datos de contenido publicado de Crossref

**● Excluir del Reporte de Similitud**

- Material bibliográfico
- Coincidencia baja (menos de 20 palabras)
- Material citado
- Bloques de texto excluidos manualmente

## PRESENTACIÓN

SEÑOR DECANO DE LA FACULTAD DE CIENCIAS QUÍMICAS, FÍSICAS Y MATEMÁTICAS DE LA UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO.

SEÑOR DIRECTOR DE LA ESCUELA PROFESIONAL DE MATEMÁTICA.

SEÑORES DOCENTES MIEMBROS DEL JURADO.

En cumplimiento con las normas y reglamento de Grados y Títulos establecidos por la escuela profesional de Matemática con Mención en Estadística, presento a vuestra consideración el presente trabajo de tesis intitulado “FACTORES QUE INFLUYEN EN LA DESERCIÓN UNIVERSITARIA DE LA ESCUELA PROFESIONAL DE MATEMÁTICA DE LA UNSAAC CON EL MODELO DE REGRESIÓN LOGÍSTICA, CUSCO 2022-2023”, con el fin de Optar el Título Profesional de Licenciado en Matemática Mención Estadística.

En este trabajo de investigación presentamos la importancia de emplear técnicas predictivas, como la minería de datos, para visualizar los motivos detrás de la deserción estudiantil. Estas técnicas posibilitan el análisis de patrones estadísticos, proporcionando a las autoridades institucionales datos fundamentados para la toma de decisiones con el fin de disminuir la tasa de deserción universitaria. Se centra en investigar los factores que influyen en la deserción de los estudiantes de la Escuela Profesional de Matemática durante el periodo 2022-2023 mediante el uso del análisis de regresión logística binaria.

El estudio de investigación se realizó para identificar los factores académicos y socioeconómicos que influyeron en la deserción universitaria de los estudiantes de la escuela profesional de Matemática, utilizando el modelo de regresión logística.

Finalmente espero que el presente trabajo de investigación trate de explicar con claridad los conceptos, definiciones desarrolladas y sirvan como guía en la formación profesional.

Atentamente,

Br. Franklin Cuchillo Valverde

## DEDICATORIA

Doy gracias al SEÑOR, porque él es bueno, su gran amor perdura para siempre, doy gracias al Dios de dioses, al único. (Salmos 136:1-2), por su gran amor incondicional que me demuestra día a día.

Agradezco a mis, queridos padres: Sra. Martha Valverde y Sr. Ángel Cuchillo, por su apoyo incondicional y su amor; debo resaltar de manera muy especial a mi querida y amada madre Martha por sus consejos y sus sabias palabras siempre motivadoras de todas aquellas veces en las que me dijo que no me rindiera y que todo es posible con Dios y es mi ejemplo a seguir.

A mis queridos hermanos: Alexander, Tania, Jackelyne, Ángel, Carina, Edwin y Arístides por sus consejos y apoyo incondicional, que son la razón de mi existencia por nuestras venas fluye la misma sangre.

## AGRADECIMIENTO

A Dios, por haberme dado la vida, sabiduría, salud y por siempre estar a mi lado en los momentos difíciles y haberme permitido concluir mi etapa universitaria y ahora el presente trabajo de investigación.

A la Escuela Profesional de Matemática de la Facultad de Ciencias Químicas, Físicas y Matemáticas de la Universidad Nacional de San Antonio Abad del Cusco y a mis docentes del Departamento Académico de Matemáticas y Estadística por compartir sus conocimientos en las diferentes asignaturas, y absolviendo toda clase de inquietudes, y así permitiéndome formarme Profesionalmente, y por las experiencias durante mi formación Profesional.

Agradecimiento total, mi asesora Dra. Nelly María Salazar Peña, por brindarme su orientación sus conocimientos trascendentales, comprensión, guía incondicional y su motivación a la rigurosa investigación, que fueron muy valioso para la realización del presente trabajo de investigación.

Mi agradecimiento también va dirigido al Dr. Epifanio Puma Huañec, Director de la escuela profesional de Matemática, por darme el permiso y la facilidad para aplicar el cuestionario dentro de la Escuela de Matemática, y llevar adelante esta investigación.

Y por último agradezco a mis compañeros de la Escuela Profesional de Matemática, quienes colaboraron con sus respuestas sinceras y muy dispuestos con el cuestionario, y que todo esto sea para el bien de nuestra Escuela.

## ÍNDICE

PRESENTACIÓN.....	ii
DEDICATORIA .....	iii
AGRADECIMIENTO.....	iv
ÍNDICE .....	v
RESUMEN.....	viii
ABSTRACT.....	ix
INTRODUCCIÓN .....	x
CAPITULO I.....	1
PLANTEAMIENTO DEL PROBLEMA.....	1
1.1.    Situación problemática.....	1
1.2.    Formulación del problema .....	4
1.2.1.    Problema general.....	4
1.2.2.    Problemas específicos .....	4
1.3.    Justificación de la investigación.....	4
1.4.    Objetivos de la investigación .....	5
1.4.1.    Objetivo general .....	5
1.4.2.    Objetivos específicos.....	5
CAPITULO II .....	7
MARCO TEÓRICO CONCEPTUAL .....	7
2.1.    Bases teóricas .....	7
2.1.1.    Modelo lineal general.....	7
2.1.2.    Regresión logística .....	8
2.1.2.1.    Ajuste del modelo de regresión logística simple.....	13
2.1.2.2.    Prueba de hipótesis para los parámetros.....	15
2.1.2.3.    Estimación por intervalo de confianza. ....	18
2.1.2.4.    Prueba de Hosmer – Lemeshow.....	19
2.1.2.5.    Odd ratio.....	20
2.1.2.6.    Tasa de clasificación correcta .....	21
2.1.2.7.    Índice de youden .....	22
2.1.3.    Deserción.....	222
2.1.3.1.    Modelos de deserción.....	23
2.1.3.2.    Teorías de deserción.....	26
2.1.3.3.    Perfil del desertor .....	28
2.1.3.4.    Actores que intervienen en la deserción.....	29
2.1.4.    Consecuencias de la deserción .....	30

2.1.4.1.    Estrategias para la prevención de la deserción .....	33
2.2.    Marco Conceptual .....	38
2.3.    Antecedentes Empíricos de la Investigación (estado del arte) .....	39
2.3.1.    Antecedentes internacionales .....	39
2.3.2.    Antecedentes nacionales .....	41
2.3.3.    Antecedentes locales .....	43
CAPITULO III .....	44
HIPÓTESIS Y VARIABLES .....	44
3.1.    Hipótesis .....	44
3.1.1.    Hipótesis general .....	44
3.1.2.    Hipótesis específicas .....	44
3.2.    Identificación de variables e indicadores .....	45
3.2.1.    Variable independiente .....	45
3.2.2.    Variable dependiente .....	45
3.3.    Operacionalización de variables .....	46
CAPITULO IV .....	48
METODOLOGÍA .....	48
4.1.    Ámbito de Estudio: localización política y geográfica .....	48
4.1.1.    Área de estudio .....	48
4.1.2.    Lugar de ejecución de la investigación .....	48
4.2.    Tipo y nivel de investigación .....	48
4.3.    Unidad de análisis .....	49
4.4.    Población de estudio .....	49
4.5.    Plan de recolección de datos .....	50
4.6.    Tamaño de muestra .....	51
4.7.    Técnicas de selección de muestra .....	52
4.8.    Técnicas de análisis e interpretación de la información .....	52
CAPITULO V .....	53
RESULTADOS .....	53
DISCUSIÓN DE RESULTADOS .....	68
CONCLUSIONES .....	70
RECOMENDACIONES .....	71
BIBLIOGRAFÍA .....	72
VALIDACIÓN DE INSTRUMENTOS .....	78



## LISTA DE TABLAS

<b>Tabla N°1:</b> Descripción de la muestra .....	<b>53</b>
<b>Tabla N°2:</b> Nivel de valoración de las dimensiones .....	54
<b>Tabla N°3:</b> Estimación del modelo de regresión logística.....	55
<b>Tabla N°4:</b> Selección de Variables .....	57
<b>Tabla N°5:</b> Modelo de regresión logística final .....	57
<b>Tabla N°6:</b> Odd ration .....	58
<b>Tabla N°7:</b> Análís del factor académico .....	59
<b>Tabla N°8:</b> Análís del factor institucional .....	60
<b>Tabla N°9:</b> Análís del factor individual .....	61
<b>Tabla N°10:</b> Análís del factor socioeconómico .....	62
<b>Tabla N°11:</b> Matriz de confusión .....	63
<b>Tabla N°12:</b> Nueva matriz de confusión .....	67

## LISTA DE GRAFICAS

<b>Grafica N°1:</b> Estimación del modelo de regresión logística .....	<b>56</b>
<b>Grafica N°2:</b> Modelo de regresión logística final .....	58
<b>Grafica N°3:</b> Análís del factor académico .....	59
<b>Grafica N°4:</b> Análís del factor institucional .....	60
<b>Grafica N°5:</b> Análís del factor individual .....	61
<b>Grafica N°6:</b> Curva ROC para el modelo logístico .....	65

## RESUMEN

El objetivo de este trabajo de investigación es determinar los factores que influyen significativamente en la deserción universitaria de los estudiantes de la escuela profesional de matemática de la Universidad Nacional San Antonio Abad del Cusco, con el modelo de regresión logística, Cusco 2022-2023. Se consideran 4 factores como son el factor académico, factor institucional, factor individual y factor socioeconómico. Esto permite medir y evaluar si dichos factores influyen en la deserción universitaria. El estudio es de nivel básico, con un alcance predictivo, enfoque cuantitativo y diseño no experimental transversal. La población estuvo compuesta por 244 estudiantes matriculados durante los semestres 2022-I al 2023-II y la muestra son de 150 estudiantes universitarios, los resultados revelan que el Factor Académico y el Factor Individual son significativos, que impactan la deserción universitaria. Específicamente, se encontró que los estudiantes con baja motivación académica y una preparación universitaria insuficiente tienen una mayor probabilidad de abandonar sus estudios. Asimismo, el reconocimiento y prestigio de la carrera influyen considerablemente en la decisión de continuar o no con los estudios universitarios.

**Palabras Claves:** Deserción universitaria, Modelo de Regresión Logística Binaria y Índice de Youden.

## ABSTRACT

The objective of this research work is to determine the factors that significantly influence the university dropout of students of the professional school of mathematics of the Universidad Nacional San Antonio Abad del Cusco, with the logistic regression model, Cusco 2022-2023. Four factors are considered such as the academic factor, institutional factor, individual factor and socioeconomic factor. This allows us to measure and evaluate whether these factors influence university dropout. The study is basic level, with a predictive scope, quantitative approach and non-experimental cross-sectional design. The population consisted of 244 students enrolled during the semester 2022-I to 2023-II and the sample is 150 university students. The results reveal that the Academic Factor and the Individual Factor are significant, which impact university dropout. Specifically, it was found that students with low academic motivation and insufficient university preparation are more likely to drop out. Likewise, the recognition and prestige of the career have a significant influence on the decision to continue or not to continue with university studies.

Key words: University dropout, Logistic regression and Indic of Youden.

## INTRODUCCIÓN

La deserción universitaria se constituye como un desafío multidimensional de alcance global para las instituciones de educación superior, incidiendo significativamente en la trayectoria académica y el desarrollo integral de los estudiantes. En el contexto específico de la Universidad Nacional de San Antonio Abad del Cusco (UNSAAC), particularmente en la Escuela Profesional de Matemática.

La presente investigación se propone identificar y analizar de manera exhaustiva los factores determinantes en la deserción de los estudiantes de la Escuela Profesional de Matemática de la UNSAAC durante el bienio 2022-2023. Mediante la aplicación de un riguroso análisis de regresión logística binaria, se examinarán indicadores académicos, institucionales, individuales y socioeconómicos asociados al fenómeno deserción de los estudios universitarios. La comprensión profunda de estos factores resulta fundamental para la formulación e implementación de estrategias de retención estudiantil eficaces y el diseño de políticas institucionales integrales. Este estudio no solo contribuirá al corpus de conocimientos locales sobre la deserción universitaria, sino que también aportará conocimientos valiosos para la comunidad académica global, considerando la universalidad del fenómeno y la heterogeneidad de sus manifestaciones en diversos contextos.

El presente trabajo de investigación está estructurado en capítulos de la forma siguiente:

CAPITULO I, en este capítulo, donde se describe de forma general, se explica los fundamentos para la realización de la tesis, así como el planteamiento del problema, situación problemática, formulación del problema, objetivos de la investigación, justificación de la investigación y limitaciones de la investigación.

CAPITULO II, en el capítulo se describe los Antecedentes, bases teóricas, Regresión Logística, Deserción, Marco Conceptual.

CAPITULO III, en el capítulo doy a conocer sobre la Hipótesis, Identificación de Variables e Indicadores y la Operacionalización de variables.

CAPITULO IV, en este capítulo se describe la metodología. Ámbito de estudio: Localización Política y Geográfica. Tipo y nivel de investigación. Unidad de análisis. Población, tamaño de muestra. Técnicas de selección de muestra. Técnicas de análisis e interpretación de la información.

CAPITULO V, en este capítulo se presenta los Resultados de la investigación.

Finalmente, se describe la conclusión, recomendaciones, bibliografía y anexos (de los datos y/o encuestas realizadas a los alumnos).

## CAPITULO I

### PLANTEAMIENTO DEL PROBLEMA

#### 1.1. Situación Problemática

La deserción universitaria, según la UNESCO (2006), se entiende como el abandono de los estudios superiores, ya sea por elección personal o por circunstancias externas que impactan al estudiante. Este proceso implica la interrupción de la carrera académica en la que el estudiante inicialmente se inscribe. Se resalta de esta definición que la deserción universitaria no se limita a ser simplemente el abandono de un programa académico en el que un estudiante se ha inscrito; en cambio, se reconoce como un proceso integral que implica diversas etapas, fases, elementos y factores. Este enfoque demanda una observación detallada y reflexiva para comprender de manera completa la complejidad de la problemática asociada a la deserción estudiantil.

Según el estudio realizado por González y su equipo de investigadores (2018), la deserción universitaria ha sido objeto de análisis en el ámbito académico, definiéndola como el fenómeno que abarca a aquellos estudiantes que, tras su matriculación en un año académico, deciden abandonar sus estudios en el año siguiente. Esta problemática puede manifestarse tanto en el abandono de la institución antes de completar el programa educativo como posteriormente a la finalización del mismo. Este tema ha experimentado una evolución integral en los últimos años, evidenciando la necesidad de comprender sus causas y efectos en el contexto educativo contemporáneo.

En la actualidad, el problema de la deserción representa un desafío significativo para las instituciones educativas a nivel global. Este problema se caracteriza por su impacto abarcador, afectando múltiples dimensiones en la vida de los individuos, incluyendo el ámbito familiar, la interacción social y el panorama laboral. La

problemática de la deserción universitaria trasciende fronteras, impactando no solo a nivel local sino también a instituciones educativas en el extranjero. Este fenómeno está generando un aumento alarmante en las estadísticas de estudiantes que optan por abandonar sus estudios durante el primer año académico. Esta tendencia, que ha evolucionado de manera significativa en los últimos años, plantea desafíos importantes para las universidades y pone de manifiesto la necesidad de abordar este tema con estrategias efectivas y en constante evolución.

En los Estados Unidos, la deserción universitaria de pregrado se sitúa en un preocupante 40% según datos generales. En contraste, en naciones de la Unión Europea como el Reino Unido, Noruega y Francia, las tasas de abandono universitario muestran cifras más moderadas, alcanzando el 16%, 17% y 24%, respectivamente. Estas disparidades en las estadísticas subrayan la importancia de abordar la problemática de la deserción estudiantil desde una perspectiva internacional. No obstante, Italia presenta la tasa más elevada de deserción, alcanzando el 33%, seguido de cerca por los Países Bajos con un 31%. (Rudin, 2019)

Se observa que el 50% de las personas de entre 25 y 29 años en la región de América Latina no lograron concluir sus estudios universitarios, a pesar de haberse matriculado. Esta situación, que afecta a una parte significativa de la población, no siempre es plenamente reconocida por las autoridades universitarias. En el caso específico de Colombia, se registró una tasa de deserción estudiantil del 31%, mientras que en Chile esta cifra fue del 21% durante el año 2019. (Behr y otros, 2020).

Según Dávila y colaboradores (2022), quienes hacen referencia al II Informe Bienal sobre la realidad universitaria en Perú, elaborado por SUNEDU en 2020, se observa que durante el período comprendido entre 2012 y 2018, la deserción de estudiantes universitarios a nivel nacional varió entre el 15.8% y el 17.6%, dependiendo

de la región geográfica. Estos informes ofrecen un análisis más detallado, resaltando que, en la selva, la tasa de deserción universitaria alcanzó el 24.6%, en la costa fue del 24%, y en la sierra fue del 18.2%.

Según los informes de la Dirección de Sistemas de Información, Unidad de Estadística de la Universidad Nacional de San Antonio Abad del Cusco (UNSAAC) para los años 2016 y 2017, se evidenció que aproximadamente el 27% y 29.69%, respectivamente, de los estudiantes que ingresaron a la institución abandonaron sus estudios universitarios. Es relevante destacar que, en la Escuela Profesional de Matemática de la misma universidad, estas cifras fueron aún más preocupantes, alcanzando el 64% y el 68%. El 62% de los estudiantes respectivamente decidió abandonar sus estudios. Estos datos ilustran las tasas de deserción en la UNSAAC y específicamente en la mencionada escuela. Es relevante destacar la importancia de emplear técnicas predictivas, como la minería de datos, para visualizar los motivos detrás de la deserción estudiantil. Estas técnicas posibilitan el análisis de patrones estadísticos, proporcionando a las autoridades institucionales datos fundamentados para la toma de decisiones con el fin de disminuir la deserción universitaria. Se centra en investigar los factores que influyen en la deserción de los estudiantes de la Escuela Profesional de Matemática durante el periodo 2022-2023 mediante el uso del análisis de regresión logística binaria.

En tal sentido se plantea las siguientes preguntas problemáticas:



## **1.2. Formulación del Problema**

### **1.2.1. Problema General**

¿Cuáles son los factores que influyen en la deserción universitaria de los estudiantes de la escuela profesional de matemática de la UNSAAC, con el modelo de regresión logística, Cusco 2022-2023?

### **1.2.2. Problemas Específicos**

- ¿Cuáles son los indicadores académicos que influyen en la deserción universitaria de los estudiantes de la escuela profesional de matemática de la UNSAAC, con el modelo de regresión logística, Cusco 2022-2023?
- ¿Cuáles son los indicadores institucionales que influyen en la deserción universitaria de los estudiantes de la escuela profesional de matemática de la UNSAAC, con el modelo de regresión logística, Cusco 2022-2023?
- ¿Cuáles son los indicadores individuales que influyen en la deserción universitaria de los estudiantes de la escuela profesional de matemática de la UNSAAC, con el modelo de regresión logística, Cusco 2022-2023?
- ¿Cuáles son los indicadores socioeconómicos que influyen en la deserción universitaria de los estudiantes de la escuela profesional de matemática de la UNSAAC, con el modelo de regresión logística, Cusco 2022-2023?

## **1.3. Justificación de la Investigación**

El problema de la deserción universitaria es de gran relevancia en el ámbito académico y en la sociedad contemporánea. Nos proporciona la oportunidad de identificar y comprender los factores determinantes que conducen a los jóvenes a abandonar sus estudios universitarios. Además, esto plantea desafíos tanto a nivel personal como familiar para los estudiantes y, al mismo tiempo, obstaculiza el desarrollo social y

económico de nuestra región. Por lo tanto, en esta tesis titulada “FACTORES QUE INFLUYEN EN LA DESERCIÓN UNIVERSITARIA DE LA ESCUELA PROFESIONAL DE MATEMÁTICA DE LA UNSAAC, CON EL MODELO DE REGRESIÓN LOGÍSTICA, CUSCO 2022-2023” resulta de suma importancia analizar los factores determinantes que ejercen una influencia significativa en la decisión de abandono por parte de los estudiantes universitarios. La relevancia de estos descubrimientos se revela como esencial para las instituciones académicas, ofreciéndoles una brújula clara para discernir los elementos críticos que subyacen a la problemática de la deserción estudiantil. Estos resultados, a su vez, facilitarán la creación de estrategias concretas orientadas a afrontar con eficacia este desafío. Por último, el presente estudio permitirá realizar investigaciones integrales sobre la variable investigada, incluida la elaboración y utilización de instrumentos como cuestionarios y programas informáticos estadísticos para la recogida de datos.

#### **1.4. Objetivos de la Investigación**

##### **1.4.1. Objetivo General**

Determinar los factores que influyen significativamente en la deserción universitaria de los estudiantes de la escuela profesional de matemática de la UNSAAC, con el modelo de regresión logística, Cusco 2022-2023.

##### **1.4.2. Objetivos Específicos**

- Determinar los indicadores académicos que influyen significativamente en la deserción universitaria de los estudiantes de la escuela profesional de matemática de la UNSAAC, con el modelo de regresión logística, Cusco 2022-2023.

- Determinar los indicadores institucionales que influyen significativamente en la deserción universitaria de los estudiantes de la escuela profesional de matemática de la UNSAAC, con el modelo de regresión logística, Cusco 2022-2023.
- Determinar los indicadores individuales que influyen significativamente en la deserción universitaria de los estudiantes de la escuela profesional de matemática de la UNSAAC, con el modelo de regresión logística, Cusco 2022-2023.
- Determinar los indicadores socioeconómicos que influyen significativamente en la deserción universitaria de los estudiantes de la escuela profesional de matemática de la UNSAAC, con el modelo de regresión logística, Cusco 2022-2023.

### **1.4.3 Limitaciones de la Investigación**

Una de las limitaciones que quizá causó mayor dificultad fue la recolección de datos de aquellos estudiantes desertores, al ya no estar en la escuela, como también la falta de información del centro de cómputo UNSAAC, una información más específica como es el caso de la deserción y porque no también manejar datos estadísticos sobre este tema. y mucho menos la Escuela de Matemática cuenta con esta información, es por ello que me vi en la necesidad de solicitar información de los estudiantes, al centro de cómputo, para así yo poder recabar información con certeza, datos sobre deserción en la Escuela de Matemática para poder utilizar en esta investigación, siempre y cuando con absoluta discreción y no sean revelados los datos personales de aquellos estudiantes.

## CAPITULO II

### MARCO TEÓRICO CONCEPTUAL

#### 2.1. Bases Teóricas

##### 2.1.1. Modelo Lineal General

El Modelo Lineal General surge por la necesidad de expresar en forma cuantitativa relaciones entre un conjunto de variables, en la que una de ellas es llamada variable respuesta o dependiente siendo las otras variables explicativas o independientes.

Sea  $Y$  una variable aleatoria cuya función de distribución de probabilidad pertenece a una familia de distribuciones de probabilidad  $H$  y es explicada por un conjunto de variables  $X_1, X_2, X_3, \dots, X_k$  las cuales son fijadas antes de conocer  $Y$ . La esperanza condicional de  $Y$  está dada por:

$$E(Y/X_1, X_2, X_3, \dots, X_k) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k \quad (1)$$

Al extraer una muestra aleatoria de tamaño  $n\{(y_i, x_{i1}, \dots, x_{ik}): i = 1, 2, \dots, n\}$ , de una población en la cual la variable respuesta  $Y$ , y las variables independientes  $X_1, X_2, X_3, \dots, X_k$ , se relacionan linealmente, cada observación de la muestra puede ser expresada como:

$$y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \varepsilon_i \quad (2)$$

En la ecuación (2), el término  $\varepsilon_i$  es una perturbación aleatoria no observable denominada error aleatorio, la cual tiene esperanza cero, varianza  $\sigma^2$  (constante); y dos errores cualesquiera  $\varepsilon_i$  y  $\varepsilon_{i'}$ ,  $\forall i \neq i'$ , son incorrelacionados entre sí.

Utilizando la notación matricial, la ecuación se puede expresar como:

$$Y = X\beta + \varepsilon \quad (3)$$

Donde,  $Y' = (Y_1, Y_2, \dots, Y_n)$  es un vector de variables aleatorias observables, denominado vector respuesta de orden  $n$ ;  $X$  es la matriz de variables independientes de orden  $n \times (k + 1)$  el vector de parámetros desconocidos de orden  $(k + 1)$  y  $\varepsilon$  el vector de errores aleatorios de orden  $n$ .

El vector de respuestas  $Y$  de la expresión está formado por dos componentes, una sistemática y otra aleatoria. La primera componente constituida por la combinación lineal  $X\beta$ , predictor lineal, el cual está representado como:

$$n = X\beta \quad (4)$$

La segunda componente, formada por el vector aleatorio  $Y$ , con elementos independientes entre sí, caracterizada por una distribución  $h \in H$  con vector de esperanzas  $\mu$  y matriz de covarianza  $\sigma^2 I$ .

Por otro lado, calculando la esperanza de  $Y$  en (3) se tiene que:

$$E(Y) = X\beta$$

Una característica distintiva del modelo lineal general, es que la variable respuesta  $Y$  está medida en escala numérica, mientras que las variables independientes pueden ser numéricas o categóricas y además son independientes entre sí. (Gonzales King-Keé, 2001)

### 2.1.2. Regresión Logística

Los modelos de regresión logística son herramientas que permiten explicar el comportamiento de una variable respuesta discreta (binaria o con más de dos categorías) a través de una o varias variables independientes explicativas de naturaleza cuantitativa y/o cualitativa. Según el tipo de variable respuesta estaremos hablando de regresión

logística binaria (variable dependiente con 2 categorías), o de regresión logística multinomial (variable dependiente con más de 2 categorías), pudiendo ser esta última de respuesta nominal u ordinal. Los modelos de respuesta discreta son un caso particular de los modelos lineales generalizados formulados por Nelder y Wedderburn en 1972, al igual que los modelos de regresión lineal o el análisis de la varianza.

El enfoque principal de los modelos de Regresión Lineal radica en la exploración de situaciones donde la variable dependiente  $Y$  se define de forma cuantitativa. Sin embargo, nos encontramos con frecuencia con situaciones en las que la variable dependiente adquiere una característica cualitativa. Por ejemplo, podemos estar interesados en analizar el color de ojos en una muestra de individuos, lo que resulta en diversas categorías como marrón, azul o negro. En tales escenarios, el modelo de Regresión Lineal no resulta adecuado. No obstante, en ciertos contextos, su aplicación puede ser beneficiosa para investigar los efectos de múltiples variables explicativas en una variable dependiente continua. Para abordar situaciones en las que la variable dependiente es cualitativa, se han desarrollado otros enfoques, entre los cuales destaca el modelo de Regresión Logística. Este enfoque, sumamente efectivo, ofrece un medio sólido para comprender y evaluar la relación entre variables explicativas y una variable dependiente categorizada. (Ancco Cayllahua, 2017)

Se utiliza la Regresión Logística como un método estadístico para estimar la probabilidad de que un grupo de personas pertenezca a una categoría específica que sea relevante. En este enfoque, la variable dependiente  $Y$  se concibe como binaria, denotando la probabilidad de que cada observación pertenezcan a una de dos clases posibles (0 o 1). La Regresión Logística establece un vínculo entre la variable objetivo y un conjunto de variables predictoras que ejercen influencia sobre ella, las cuales pueden adoptar tanto naturaleza cualitativa como cuantitativa. Es importante destacar que estas variables

predictoras deben ser independientes entre sí, es decir, no deben estar correlacionadas. Esto asegura que el modelo pueda estimar con precisión la probabilidad de pertenencia a una clase en función de las variables predictoras seleccionadas. (Ancco Cayllahua, 2017)

Se introduce una variable aleatoria binaria que se rige por una distribución de Bernoulli, presentando dos resultados posibles: 1 o 0, con una probabilidad de  $\pi(x)$  y  $1 - \pi(x)$  respectivamente. En este contexto,  $\pi(x)$  representa la probabilidad condicional de que  $Y$  asuma el valor de 1 dado  $x$ . Para modelar esta variable de respuesta binaria, se recurre a la distribución logística, la cual se formula de la siguiente manera:

$$\pi(x) = \frac{e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}} \quad (5)$$

Donde  $x = (x_1, x_2, \dots, x_p) \in R^p$  se refiere a un vector que representa  $p$  variables explicativas y  $\beta_0, \beta_1, \dots, \beta_p$  son parámetros desconocidos.

Cumplir con los supuestos de linealidad y la naturaleza continua de las variables suele resultar difícil al trabajar con modelos de Regresión Lineal. La Regresión Logística presenta una ventaja al no requerir estrictamente que las variables independientes sigan una distribución específica. Esto implica que las variables predictoras pueden ser de cualquier tipo, ya sean continuas o cualitativas, y no es necesario que sigan una distribución normal. Esta flexibilidad hace que la Regresión Logística sea una herramienta más adaptable y poderosa para modelar relaciones entre variables en comparación con la Regresión Lineal. (Hosmer & Lemeshow, 2000)

La Regresión Logística proporciona una estimación de la probabilidad de que un individuo específico pertenezca a uno de dos grupos, por ejemplo, aquellos estudiantes que aprueban y aquellos que no, teniendo en cuenta. Este modelo estadístico permite calcular qué tan probable es que un individuo sea parte de un grupo en particular en

función de ciertas variables predictoras, lo que lo convierte en una herramienta valiosa para predecir la pertenencia a categorías o conjuntos diferentes en diversos contextos, como el académico o el de investigación.

En el marco de esta investigación, la variable dependiente, denotada como  $Y$ , de naturaleza binaria, adquiere el valor "0" cuando el evento en cuestión no ocurre, y "1" en caso contrario. Las variables independientes, agrupadas bajo la designación  $X$ , pueden presentar tanto características cuantitativas como cualitativas. Se representa como  $p$  la probabilidad de que la variable  $Y$  adquiera el valor "1", lo que indica la probabilidad de que se materialice el evento de interés en el marco del estudio. Dentro del contexto de los modelos de regresión, el enfoque principal radica en estimar el valor medio de la variable de respuesta, el cual está influenciado por un conjunto de variables predictoras. Estas variables regresoras, a su vez, contribuyen a determinar la tendencia central de la variable dependiente, proporcionando una representación significativa de la relación entre las variables en estudio. La noción de "media condicional" se emplea para describir este concepto, representado como  $E(Y/x)$ , donde " $Y$ " y " $X$ " denotan los valores de la variable dependiente e independiente, respectivamente. Dentro del contexto de los modelos de regresión lineal, se parte del supuesto de que la media puede ser modelada mediante una ecuación lineal que depende de " $x$ ". En estos modelos, se asume que la media puede ser expresada a través de una ecuación lineal que se fundamenta en " $x$ ":

$$E(Y/x) = \beta_0 + \beta_1 x \quad (6)$$

Esta fórmula establece que  $E(Y/x)$  puede asumir cualquier valor de  $x$ , debe encontrarse en el rango de  $-\infty$  a  $+\infty$ . Para simplificar y representar el valor medio de  $Y/x$  al aplicar la distribución logística, se empleará la notación  $\pi(x) = E(Y/x)$ . La formulación específica del modelo de Regresión Logística se desarrolla de esta manera.



$$P(y = 1) = \pi(x) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \quad (7)$$

Dentro del campo de la regresión logística, el Logit es una modificación importante de  $\pi(x)$ . Esta modificación se establece en relación con  $\pi(x)$  de la siguiente forma:

$$\begin{aligned} g(x) &= \ln \left[ \frac{\pi(x)}{1 + \pi(x)} \right] \\ &= \beta_0 + \beta_1 X \end{aligned} \quad (8)$$

La inclusión de la función  $g(x)$  en esta situación sugiere que comparte múltiples atributos característicos de la regresión lineal. Una distinción fundamental entre los modelos de regresión lineal y logística reside en la distribución condicional de la variable dependiente. En los modelos lineales, se asume que un valor de la variable dependiente puede ser expresado como  $y = E(Y/x) + \varepsilon$ , donde  $\varepsilon$  denota el error que refleja la discrepancia entre un valor y la media condicional. Una suposición importante es que  $\varepsilon$  sigue una distribución gaussiana o normal con una media de cero y una varianza constante en los distintos niveles de la variable independiente.

“Entonces, se infiere que la distribución de la variable dependiente, dado  $x$ , seguirá una distribución normal con media  $E(Y/x)$  y varianza constante. La representación de la variable dependiente, condicionada a  $x$  en un entorno binario, se puede expresar como  $y = E(Y/x) + \varepsilon$ , donde  $y$  puede tener uno de dos valores posibles. Si  $\varepsilon = 1 - \pi(x)$  con una probabilidad de  $\pi(x)$ , entonces  $Y = 1$ . Por lo tanto,  $\varepsilon$  sigue una distribución con media cero y varianza equivalente a  $\pi(x)[1 - \pi(x)]$ ”.

“En síntesis, la explicación de la distribución condicional de la variable dependiente se fundamenta en una distribución binomial, cuya probabilidad de éxito se encuentra definida por la media condicional  $\pi(x)$ ”. (Hosmer & Lemeshow, 2000)

### 2.1.2.1. *Ajuste del modelo de regresión logística simple.*

“Consideremos un conjunto de "n" atributos que son mutuamente independientes, representados por pares  $(x_i, y_i)$ , donde "i" puede variar de 1 a n. En este contexto,  $y_i$  denota que la variable de respuesta adquiere el valor "y", mientras que  $x_i$  representa el valor de la variable independiente para el i-ésimo individuo. Además, suponemos que la variable de respuesta puede ser 0 o 1, donde 1 indica la presencia de una característica particular y 0 indica lo contrario. Para ajustar el modelo de Regresión Logística, como se muestra en la ecuación (8), a un conjunto de datos, es necesario estimar los valores de  $\beta_0$  y  $\beta_1$ , los cuales son desconocidos”. (Hosmer & Lemeshow, 2000)

En el ámbito de la Regresión Lineal, el método convencional para estimar los parámetros desconocidos es el Método de Mínimos Cuadrados (MMC). Sin embargo, a diferencia de la Regresión Logística, donde el MMC presenta ciertas propiedades estadísticas particulares, estas características no se conservan cuando la variable de salida es binaria, como en la Regresión Logística.

Para abordar este desafío, la estimación de parámetros en la Regresión Logística se basa en el enfoque de Máxima Verosimilitud (MV). Este método busca determinar los valores de los parámetros desconocidos que maximizan la probabilidad de que los datos observados sean el resultado obtenido. La metodología de MV comienza con la construcción de una función conocida como Función de Verosimilitud.

“Cuando clasificamos la expresión de  $\pi(x)$  obtenida en la ecuación (7) como 0 o 1, determinada por los valores arbitrarios de  $\beta = (\beta_0, \beta_1)$ , estamos indicando la probabilidad condicional de que  $Y$  sea igual a 1 dado un valor de  $x$ . Esto se denota como  $P(Y = 1/x)$ . A partir de esta observación, se infiere que  $1 - \pi(x_i)$  de la probabilidad condicional de  $Y$  determinada por un valor de  $x$  es igual a cero, es decir,  $P(Y = 0/x)$ ”.

“Para los diversos pares  $(x_i, y_i)$ , donde  $y_i = 1$ , se añade su respectiva contribución a la función de probabilidad, representada por  $\pi(x_i)$ ; por otro lado, cuando los pares están definidos por  $y_i = 0$ , la contribución correspondiente a la función de probabilidad se establece como  $1 - \pi(x_i)$ . Aquí, el término  $\pi(x_i)$  hace referencia a los valores calculados de  $\pi(x)$  evaluados en  $x_i$ . Otra manera apropiada de expresar la contribución a la función de probabilidad para el par  $x_i, y_i$  es a través de una función”. (Hosmer & Lemeshow, 2000)

$$\pi(x_i)^{y_i} = [1 - \pi(x_i)]^{1-y_i} \quad (9)$$

“Considerando la suposición de independencia entre las observaciones, avanzamos en el cálculo de la función de verosimilitud multiplicando los componentes de la ecuación (9) según el siguiente procedimiento metodológico”:

$$l(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} = [1 - \pi(x_i)]^{1-y_i} \quad (10)$$

Siguiendo el método de máxima verosimilitud, se busca maximizar la ecuación (10) respecto a los valores de  $\beta$ . “Sin embargo, por conveniencia matemática, se prefiere trabajar con el logaritmo de la ecuación (10). La expresión del logaritmo de verosimilitud se define de acuerdo con la siguiente estructura”:

$$L(\beta) = \ln[l(\beta)] = \sum_{i=1}^n \{y_i \ln [\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]\} \quad (11)$$

Para hallar los valores de  $\beta$  que maximizan  $L(\beta)$ , se procede calculando la derivada de  $L(\beta)$  con respecto a  $\beta_0$  y  $\beta_1$ , y posteriormente se iguala a cero. Las ecuaciones resultantes, denominadas ecuaciones de verosimilitud, se expresan de la siguiente manera:

$$\frac{\partial L(\beta)}{\partial \beta_0} = \sum_{i=1}^n [y_i - \pi(x_i)] = 0 \quad (12)$$

$$\frac{\partial L(\beta)}{\partial \beta_1} = \sum_{i=1}^n x_i * [y_i - \pi(x_i)] = 0 \quad (13)$$

“Las estimaciones de máxima verosimilitud para los coeficientes  $\beta$ , obtenidas al resolver las ecuaciones (12) y (13), se denotan como  $\hat{\beta}$ . Este proceso de obtención no es trivial, requiriendo la aplicación de métodos numéricos, siendo el método de Newton-Raphson uno de los más destacados”. (Hosmer & Lemeshow, 2000)

### **2.1.2.2. Prueba de hipótesis para los parámetros.**

“Después de calcular los parámetros del modelo, se enfoca en determinar la importancia relativa de las variables incluidas en él. Este procedimiento implica llevar a cabo pruebas de hipótesis para comprender la naturaleza de la relación entre la variable dependiente y las variables independientes”. (Hosmer & Lemeshow, 2000)

Dentro del marco de la Regresión Lineal, existe un procedimiento específico para analizar el impacto de las variables, el cual puede ser adaptado para su aplicación en la Regresión Logística. Este procedimiento implica evaluar la significancia de los coeficientes, seguido por la construcción de una matriz de varianzas. “Esta matriz, organizada meticulosamente, descompone la suma de cuadrados en relación con su media en dos componentes distintos: SSR y SSE. La fórmula para SSE (Suma de Cuadrados del Error) se define de la siguiente manera”:

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (14)$$

“Partiendo del modelo sin la variable de respuesta y al considerar  $\beta_0$ , donde  $\hat{\beta}_0$  y la estimación de  $\beta_0$  es  $\bar{y}$ , siendo igual a la media de la variable dependiente, SSE es equivalente a la Varianza Total. Al introducir la variable independiente en el modelo, cualquier disminución en la suma de errores cuadráticos (SSE) se atribuye al hecho de que el coeficiente  $\beta_0$  asociado a la variable independiente es distinto de cero. La variación

en el valor de SSE se debe a la variabilidad de la regresión, conocida como suma de cuadrados de la regresión (SSR), y se expresa de esta manera”.

$$SSR = [ \sum_{i=1}^n (y_i - \bar{y}_i)^2 ] - [ \sum_{i=1}^n (y_i - \hat{y}_i)^2 ] \quad (15)$$

En los contextos de Regresión Lineal, se persigue la evaluación de la significancia de la suma de cuadrados de la regresión (SSR). Un SSR elevado denota la importancia de la variable independiente en el modelo, mientras que un valor menor sugiere una contribución menos relevante de dicha variable para predecir la variable dependiente. (Hosmer & Lemeshow, 2000)

“La comparación entre los valores observados de la variable dependiente y las predicciones derivadas de modelos que incorporan la variable independiente constituye el principio esencial tanto en la Regresión Lineal como en la Regresión Logística. La determinación de la importancia del coeficiente  $\beta_0$  de la variable dependiente se realiza mediante la prueba de razón de verosimilitud, cuyo cálculo se lleva a cabo con la siguiente fórmula”.

$$D = -2 \sum_{i=1}^n \left[ y_i \ln \left( \frac{\hat{\pi}_i}{y_i} \right) + (1 - y_i) \ln \left( \frac{1 - \hat{\pi}_i}{1 - y_i} \right) \right] \quad \text{Donde } \hat{\pi}_i = \hat{\pi}(x_i). \quad (16)$$

“Dentro de la Regresión Logística, la estadística D en la ecuación (16) se conoce como la Desviación. Esta función guarda similitudes con la suma residual de cuadrados (SSE) en la Regresión Lineal. Por lo tanto, las ecuaciones descritas en (16), si se aplican en el contexto de la Regresión Lineal, se refieren principalmente a SSE”.

Se puede expresar este valor estadístico de la siguiente manera:

$$D = -2 \ln (\text{Verosimilitud del modelo ajustado}) \quad (17)$$

“Se puede comparar el resultado de D con y sin la variable independiente incluida en el modelo para determinar su importancia. Se puede calcular el cambio en el valor de D al agregar o quitar la variable regresora del modelo de la siguiente manera”.

$$G = D (\text{Modelo sin variable}) - D (\text{Modelo con la variable}) \quad (18)$$

En la regresión logística, el valor de G cumple una función similar al numerador en la prueba F parcial de la regresión lineal. La estructura que expresa la estadística G es la siguiente.

$$G = -2 \ln \left[ \frac{\text{Verosimilitud sin la variable}}{\text{Verosimilitud con la variable}} \right] \quad (19)$$

En situaciones donde una variable independiente no forma parte del modelo, es relativamente simple demostrar que su contribución a la ecuación se anula, el estimador de máximo verosimilitud de  $\beta_0$  es equivalente a  $\ln(n_1/n_0)$  donde  $n_1 = \sum y_i$ ,  $n_0 = \sum(1 - y_i)$  Cuando el valor predicho es constante, igual a  $n_1/n$ , bajo estas suposiciones, el valor de G se presenta de la siguiente manera:

$$G = 2 \ln \left[ \frac{\binom{n_1}{n} \binom{n_0}{n}^{n_0}}{\prod_{i=1}^n \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{(1-y_i)}} \right] \quad (20)$$

$$G = 2 \left\{ \sum_{i=1}^n [y_i \ln(\hat{\pi}_i) + (1 - y_i) \ln(1 - \hat{\pi}_i)] - [n_1 \ln(n_1) + n_0 \ln(n_0) - n \ln(n)] \right\} \quad (21)$$

(Hosmer & Lemeshow, 2000)

“Además, existen otras pruebas, como la prueba de Wald y la prueba de Score, que poseen igual importancia estadística. Los supuestos empleados en estas pruebas son precisamente los mismos que los de la prueba de razón de verosimilitud”.

“La prueba de Wald implica comparar la estimación de máxima verosimilitud de  $\hat{\beta}_1$  con la estimación de su error estándar. La proporción resultante, bajo la suposición de

que  $\beta_1$  es igual a cero, sigue una distribución normal estándar. La expresión del estadístico de Wald para los modelos de Regresión Logística se presenta de la siguiente manera”.

$$W = \frac{\widehat{\beta}_1}{SE(\widehat{\beta}_1)} \quad (22)$$

W sigue una distribución normal estándar.

El test de Score se basa en la teoría de la distribución probabilística del logaritmo de verosimilitud. Se trata de una prueba multivariante que implica cálculos matriciales en su totalidad. Cuando se aplica de manera univariante, este test se fundamenta en la distribución condicional de la derivada de la ecuación (13).

“En este escenario, podemos desarrollar una expresión para el test de Score. Dicho test emplea el valor de la ecuación (13) utilizando  $\beta_0 = \ln \left( \frac{n_1}{n_0} \right)$  y  $\beta_1 = 0$ . Bajo estos parámetros,  $\widehat{\pi} = \frac{n_1}{n_0} = \bar{y}$ . Se puede demostrar que la varianza estimada es  $\sum_{i=1}^n x_i(y_i - \bar{y})$ .

El estadístico de prueba para el test de Score es”:

$$ST = \frac{\sum_{i=1}^n x_i(y_i - \bar{y})}{\sqrt{\bar{y}(1-\bar{y}) \sum_{i=1}^n (x_i - \bar{x})^2}} \quad (23)$$

### 2.1.2.3. Estimación por intervalo de confianza.

“El estimador obtenido mediante intervalos de confianza para el valor de  $\beta_1$  y el intercepto se basa en la prueba de Wald. Por lo tanto, el intervalo de confianza al 100(1- $\alpha$ ) % para  $\beta_1$  se define de la siguiente manera”.

$$\widehat{\beta}_1 \pm Z_{1-\alpha/2} SE(\widehat{\beta}_1) \quad (24)$$

Y para el intercepto es

$$\widehat{\beta}_0 \pm Z_{1-\alpha/2} SE(\widehat{\beta}_0) \quad (25)$$

“Donde  $Z_{1-\alpha/2}$  es el valor crítico correspondiente al percentil  $100(1 - \alpha) \%$  de una distribución normal estándar, y  $\widehat{SE}$  denota el estimador del error estándar de  $\widehat{\beta}_1$  calculado por el modelo. La logit, componente lineal de los modelos de Regresión Logística, se estima de la siguiente manera”:

$$\hat{g}(x) = \widehat{\beta}_0 + \widehat{\beta}_1 x \quad (26)$$

El cálculo del estimador de la varianza de  $\hat{g}(x)$  requiere sumarse. En este caso, es equivalente a:

$$\widehat{VAR}[\hat{g}(x)] = \widehat{VAR}[\widehat{\beta}_0] + x^2 \widehat{VAR}[\widehat{\beta}_1] + 2x \widehat{COV}[\widehat{\beta}_0, \widehat{\beta}_1] \quad (27)$$

“La varianza de la suma se calcula como la suma de las varianzas individuales de cada término, más la covarianza entre los elementos del intervalo de confianza al  $100(1-\alpha) \%$  según el método Wald para la logit”.

$$\hat{g}(x) \pm Z_{1-\alpha} \widehat{SE}(\hat{g}(x)) \quad (28)$$

“Donde  $\widehat{SE}(\hat{g}(x))$  representa la raíz cuadrada del estimador de la varianza proporcionado en la ecuación (27), es fundamental en este análisis”. (Hosmer & Lemeshow, 2000)

#### **2.1.2.4. Prueba de Hosmer – Lemeshow.**

El modelo de Regresión Logística, propuesto por Hosmer y Lemeshow (2000), emplea el estadístico Chi-cuadrado de Pearson para confirmar la idoneidad del ajuste. Este estadístico compara lo observado con lo esperado, evaluando si el modelo puede explicar adecuadamente los datos observados.

Hosmer–Lemeshow plantea las siguientes hipótesis:

H0: El modelo es apropiado y se ajusta a los datos



H1: El modelo no es apropiado y no se ajusta a los datos

El estadístico de prueba es:

$$\hat{C} = \sum_{j=1}^g \frac{(O_n - n_n \bar{\pi}_n)^2}{\bar{\pi}_n n_n (1 - \bar{\pi}_n)} \quad (29)$$

Donde  $O_n = \sum_{j=1}^{C_n} y_j$  representa la suma de respuestas dentro de las covariables en el n-ésimo decil, donde  $n_n$  denota la cantidad de covariables en el n-ésimo grupo,  $C_n$  es el total de sujetos en el n-ésimo grupo y

$$\bar{\pi}_k = \sum_{j=1}^{C_n} \frac{m_j \bar{\pi}_j}{n_j}$$

La representación de la estimación de la media se expresa mediante la probabilidad del valor medio estimado. Bajo la hipótesis nula, el estadístico  $\hat{C}$  sigue una distribución Chi-cuadrado con  $g-2$  grados de libertad. Se rechaza la hipótesis nula si la probabilidad del valor medio estimado es baja.

$$\hat{C} > X_{1-\alpha}^2(g-2) \quad (30)$$

y se llega a la conclusión de que el modelo no es adecuado. (Hosmer & Lemeshow, 2000)

#### 2.1.2.5. Odds Ratio (OR)

En la estadística el odd ratio es una razón de odds ósea es la probabilidad de que un evento ocurra frente a la probabilidad que no ocurra.

$$Odd Ratio = OR = \frac{Odd_A}{Odd_B} = \frac{p_A/1-p_A}{p_B/1-p_B} = \exp(\beta_i) \quad (31)$$

- Si en el análisis el resultado del  $OR > 1$ , la asociación es positiva, es decir, la presencia del factor se asocia a la mayor ocurrencia del evento, y se le considera factor de riesgo.

- Si en el análisis el resultado del  $OR < 1$ , la asociación es negativa; es decir, la presencia del factor no se asocia con la mayor ocurrencia del evento por consiguiente se considera un factor de protección.
- Si en el análisis el resultado del  $OR = 1$ , no existe asociación entre las variables, es decir que la cantidad de veces que el evento ocurra será igual con o sin la presencia del factor. (Alegre y Cahuana, 2020)

### 2.1.2.6. Tasa de Clasificación Correcta

Esta tasa permite evaluar la eficacia del modelo para clasificar nuevos individuos, ya sea en grupo de los unos o en el grupo de los ceros, se elige un punto de corte de 0.5. si el valor de la probabilidad es  $\geq 0.5$  se será clasificado como “1” si la probabilidad  $< 0.5$  será clasificado como “0”

	Probabilidades Reales		
	Si	No	Total
Probabilidades predichas	$n_{11}$	$n_{12}$	$n_{11} + n_{12}$
Total	$n_{21} + n_{22}$	$n_{12} + n_{22}$	$n$

Fuente: Alegre y Cahuana, 2020

Donde los valores de  $n_{11}$  y  $n_{22}$  son los correctamente clasificados, que es lo mismo a decir  $\frac{n_{11}+n_{22}}{n} * 100\%$ , es el porcentaje de observaciones bien clasificados por el modelo de Regresión Logística, y se espera que este porcentaje sea el más alto posible a fin de concluir que el modelo obtenido clasifica bien las probabilidades predichas, y  $n_{12}$  y  $n_{21}$  son los incorrectamente clasificados; es decir  $\frac{n_{12}+n_{21}}{n} * 100\%$ , es el porcentaje de observaciones mal clasificados.

### a. La sensibilidad

La sensibilidad del modelo se refiere a la capacidad que tiene este para detectar como positivos los casos que poseen la característica. También, dicho de otra manera, si al modelo le presentamos solo casos positivos, la sensibilidad determina la capacidad que tiene el modelo de no equivocarse.

$$\text{Sensibilidad} = \frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{Falsos negativos}} \quad (32)$$

### b. La Especificidad

La especificidad del modelo se refiere a la capacidad que tiene este para discriminar correctamente los casos que no poseen las características. Es decir, sobre un conjunto de casos que no poseen la característica. Determina en que grado no va a confundirlos con casos que poseen la característica. (Alegre y Cahuana, 2020)

$$\text{Especificidad} = \frac{\text{Verdaderos negativos}}{\text{Verdaderos negativos} + \text{Falsos positivos}} \quad (33)$$

#### 2.1.2.7. Índice Youden

El Índice de Youden se utiliza para optimizar el punto de corte del modelo, buscando maximizar tanto la sensibilidad como la especificidad. El índice de Youden se calcula mediante la siguiente formula:

$$\text{Índice de Youden} = (\text{Sensibilidad} + \text{Especificidad}) - 100 \quad (34)$$

#### 2.1.3. Deserción

Mirez, (2014) , la deserción se define como la interrupción permanente y voluntaria de los estudios, que puede manifestarse de diversas formas, como la renuncia a continuar la carrera, la salida de la institución educativa o la renuncia al sistema educativo superior en su totalidad.

Carvajal Olaya et al., (2009), sostiene que la deserción se refiere a la acción de dejar una carrera universitaria, ya sea de forma voluntaria o involuntaria, debido a circunstancias que afectan al estudiante, ya sean internas o externas.

Mairata, (2010), cuando un estudiante interrumpe su trayectoria educativa sin proseguir los estudios que había comenzado, sin inscribirse en otro programa académico o sin completar sus estudios en la universidad donde estaba inscrito, se le considera dado de baja.

Según Mirez (2014), la deserción se describe como la interrupción del compromiso que se establece tras la admisión en una institución educativa, causada por una variedad de razones, tanto del estudiante como de la institución misma. Este fenómeno tiene consecuencias significativas en términos económicos, académicos y sociales para ambas partes involucradas.

Según Narvárez Barros & Barragán Reyes,(2015) “la deserción en la educación superior se refiere al fenómeno en el que un número determinado de estudiantes abandonan el sistema de educación superior durante un periodo específico, ya sea un semestre o un año”.

“La deserción universitaria ocurre cuando un estudiante abandona su carrera universitaria de forma voluntaria o involuntaria debido a diversas circunstancias, sean estas favorables o desfavorables”. (Narvárez Barros & Barragán Reyes, 2015)

#### ***2.1.3.1. Modelos de deserción.***

Existen varios enfoques para comprender la deserción universitaria, como el psicológico, el sociológico, el económico, los enfoques organizacionales y los interaccionales.

### **Modelo psicológico.**

Desde la perspectiva psicológica, se sugiere que la elección de abandonar la universidad está influenciada por el historial de comportamiento anterior, el cual se manifiesta en actitudes y normas que fomentan el abandono temprano. (Narvárez Barros & Barragán Reyes, 2015).

Según Girón Azurdia (2014), los rasgos de personalidad juegan un papel fundamental en la distinción entre los estudiantes que culminan sus estudios y aquellos que no lo hacen. Se plantea que los comportamientos, actitudes y normas subjetivas hacia la deserción tienen un impacto adverso en las intenciones iniciales del estudiante, según este autor. Asimismo, se postula que el desempeño académico previo ejerce influencia en el rendimiento futuro al afectar la autoimagen y las percepciones de los estudiantes sobre los desafíos del aprendizaje, así como sus metas, valores y expectativas.

### **Modelos sociológicos.**

Se sostiene que la determinación del estudiante de dejar sus estudios se ve afectada por los obstáculos para integrarse al contexto universitario, resaltando la significancia del ambiente familiar. Los contratiempos familiares inciden directamente en el desempeño académico, el logro y el crecimiento intelectual del estudiante, subrayando la importancia del respaldo parental para los estudiantes. (Narvárez Barros & Barragán Reyes, 2015).

Se resalta que la determinación de abandonar los estudios está fuertemente influida por factores externos. De acuerdo con Girón Azurdia (2014), la falta de inserción del estudiante en la vida universitaria, influenciada por aspectos como el entorno familiar, las expectativas y las presiones, puede estar relacionada con la deserción temprana. Además, Spady identificó seis variables que podrían prever la deserción universitaria: aspectos académicos, sociales, socioeconómicos, de género, la calidad del programa

académico y el promedio de calificaciones semestrales son consideraciones fundamentales para la inserción.

### **Modelos económicos**

Se argumenta que la determinación de dejar los estudios está vinculada a la percepción del estudiante sobre su capacidad para hacer frente a los costos universitarios, que comprenden la matrícula, las tarifas, los cursos y los gastos administrativos. Los estudiantes que enfrentan dificultades financieras pueden tomar la decisión de abandonar la universidad debido a factores como el diseño de los programas de estudio, los descuentos ofrecidos a los estudiantes, las becas basadas en el rendimiento académico y la participación en actividades universitarias, además de los préstamos estudiantiles con tasas de interés bajas. Deberíamos considerar el impacto de estas medidas en lugar de solo pedir que el estudiante sea expulsado de la institución educativa. (Narvárez Barros & Barragán Reyes, 2015).

### **Modelos organizacionales.**

Factores importantes que inciden en la elección de un estudiante de continuar sus estudios en una institución educativa abarcan la reputación académica, la calidad educativa, la experiencia ofrecida, así como la infraestructura y la tecnología disponibles. (Narvárez Barros & Barragán Reyes, 2015).

Según lo mencionado, Berger & Milem (2014) argumentaron que la deserción universitaria está influenciada por la calidad de la experiencia social en la institución, especialmente en lo que respecta a la integración. La excelencia de la enseñanza y el grado de implicación activa de los alumnos en el proceso de aprendizaje en el aula están íntimamente ligados.

### **Modelos interaccionales.**

Se centran en el modo en que los estudiantes se relacionan con la institución universitaria y con sus pares. (Narvárez Barros & Barragán Reyes, 2015).

La adaptación de un estudiante a la vida universitaria se ve influenciada por su trasfondo familiar, sus características individuales y su experiencia previa. A medida que transcurre el tiempo en la universidad, los estudiantes logran, en parte o en su totalidad, integrarse académicamente conforme a los objetivos de su carrera y socialmente al adaptarse a sus compañeros, profesores y al sistema institucional en general. Si un estudiante no consigue esta integración, podría considerar la opción de abandonar sus estudios. Además de los aspectos universitarios. Según Girón Azurdia (2014), otros factores como la situación económica, las alternativas disponibles y las relaciones sociales externas también pueden influir en la decisión de abandonar.

#### ***2.1.3.2. Teorías de deserción.***

De acuerdo con la Teoría de Tinto, un experto en el ámbito de la educación superior, reconocido por su investigación sobre la retención estudiantil y las comunidades de aprendizaje, sostiene que la deserción en la educación superior es un fenómeno sumamente complejo. Para él, esta situación implica una variedad de enfoques y maneras de abandonar que ninguna definición probablemente pueda abarcar completamente. Según Tinto, la decisión de dejar una universidad aparentemente sencilla puede tener significados diversos y completamente diferentes para las personas involucradas o afectadas por ese comportamiento. La deserción en una institución universitaria específica se define no solo por las intenciones individuales, sino también por los procesos sociales e intelectuales que atraviesan los individuos en su búsqueda de metas académicas y personales. Aunque en estos procesos intervienen diversas fuerzas, es

también una responsabilidad esencial de los individuos alcanzar los objetivos institucionales establecidos (Tinto, 1975). La teoría de Tinto se enfoca en la integración y adaptación de los estudiantes a la vida universitaria, destacando la importancia de que los estudiantes operen según la teoría del intercambio para desarrollar una integración social y académica, evidenciada en metas y niveles de compromiso con la institución.

### **Teoría de Braxton**

**Braxton et al.** (1997) desarrollaron una teoría que divide la deserción en cinco categorías, centrándose en el análisis de variables explicativas individuales, institucionales y familiares. Estos enfoques abarcan perspectivas psicológicas, económicas, sociológicas, organizacionales e interaccionales. Según Braxton, la participación en actividades académicas es fundamental para comprender cómo la experiencia universitaria se relaciona con la retención de los estudiantes en la institución. El autor sugiere que cuanto más se implique un estudiante en actividades universitarias, menor será el riesgo de abandonar. Además, Attinasi (1986) destacó que la decisión de dejar los estudios está influenciada por las percepciones y evaluaciones que los estudiantes hacen sobre su experiencia universitaria después de inscribirse, lo que subraya la importancia de estos análisis para determinar su permanencia o abandono posterior.

### **Teoría de Bean**

Bean, en su teoría, introduce una dimensión ambiental que resalta un impacto directo en la retención estudiantil. Estos factores se derivan de las circunstancias y características personales que los estudiantes enfrentan al ingresar y durante su proceso de estudio, como la escasez de recursos económicos, así como las responsabilidades familiares y laborales. Estos elementos externos, sobre los cuales las universidades tienen un control limitado, pueden ser determinantes significativos que influyen en la decisión



de los estudiantes de abandonar sus estudios. De acuerdo con esta teoría, la elección de dejar o continuar en una universidad está estrechamente relacionada con un conjunto de variables influenciadas por el entorno, como la situación financiera, el empleo y el apoyo familiar. (Bean, 1980).

### ***2.1.3.3. Perfil del desertor***

De acuerdo con Mirez (2014), esta población puede ser identificada por "los resultados observados en varios niveles en varios programas universitarios, población que, a pesar de tener todas las oportunidades y recursos disponibles para completar su carrera universitaria y establecerse profesionalmente de manera adecuada, optan por abandonarla". las personas que abandonan las instituciones educativas, independientemente de las circunstancias, pueden manifestar algunas de las siguientes características.

**Asuntos Disciplinarios:** En general, los estudiantes que no cumplen con las normas establecidas por la institución educativa tienen más probabilidades de abandonar sus estudios. Estos estudiantes, en su mayoría, invierten más tiempo en la institución.

**Bajo nivel socioeconómico o limitaciones financieras:** Esta característica suele ser la causa principal del abandono escolar.

**Ausencia en clases:** Algunos estudiantes abandonan debido a compromisos laborales u otras responsabilidades, mientras que otros lo hacen por falta de interés en las actividades académicas.

**Problemas de salud mental:** Dificultades psicológicas que afectan el bienestar físico y emocional.

**Problemas de edad:** La rebeldía juvenil puede influir en la decisión de dejar los estudios.

**Dificultades en las relaciones interpersonales:** Los estudiantes que enfrentan dificultades para adaptarse al entorno universitario pueden tener problemas para establecer relaciones sociales y tienden a ser más introvertidos.

**Falta de participación en actividades extracurriculares:** Asistir a reuniones, participar en grupos estudiantiles y trabajar en equipo son actividades que algunos estudiantes evitan.

**Falta de interés en el aprendizaje:** Los estudiantes que abandonan la escuela a menudo carecen de una verdadera motivación para adquirir conocimientos.

**Pérdida de motivación en el trabajo o la universidad:** La elección de una carrera inadecuada es una de las principales razones detrás del abandono escolar.

#### ***2.1.3.4. Actores que intervienen en la deserción***

Según lo señalado por Mírez (2014), varios actores participan en el fenómeno de la deserción estudiantil.

**Deserción:** Estudiantes que han interrumpido su educación en una universidad.

**Padres de familia:** Padres cuyos hijos han abandonado la universidad.

**Antiguo compañero de clase:** Estudiantes que dejaron el semestre anteriormente.

**Docentes:** Profesores que han acompañado a los estudiantes desertores hasta el semestre en que se retiraron.

**Síndicos y Administradores Académicos.** Estos actores pueden ejercer influencia directa o indirecta en las decisiones de abandono estudiantil.

**Presión familiar:** En muchas ocasiones, los padres, motivados por razones financieras, instan a los estudiantes a abandonar la escuela para contribuir al ingreso familiar o buscar empleos que les permitan sostenerse económicamente. En casos de dificultades económicas familiares, los estudiantes pueden verse obligados a dejar la universidad y buscar trabajos que se ajusten a sus necesidades financieras.

**Influencia de compañeros anteriores:** Los antiguos compañeros pueden influir en las opiniones de los profesores, en la selección de materias y en la consideración de otras universidades donde los estudiantes puedan continuar sus estudios.

En lo que respecta a los profesores, directores y administradores académicos, su influencia suele manifestarse más en la calidad de la relación que mantienen con los estudiantes que en la dificultad del material académico. La falta de conexión con los asesores estudiantiles o la sensación de falta de respaldo por parte de las autoridades académicas son algunas de las razones principales por las cuales muchos estudiantes deciden abandonar sus estudios. (Mirez, 2014).

#### **2.1.4. Consecuencias de la deserción**

Según Aguirre Nativi y otros (2011), aquellos individuos que abandonan sus estudios y no se capacitan adecuadamente exhiben una productividad laboral reducida, lo que conduce, a nivel nacional, a una disminución en el crecimiento económico en general.

Así, Cuando hay áreas significativas de deserción educativa en una sociedad de manera recurrente, se perpetúan grandes disparidades sociales y económicas. Es decir, si hay menos estudiantes que abandonan la escuela y más personas se inscriben en programas educativos, aumentará el porcentaje de la población que se educa, lo que enriquece la cultura general de la sociedad.

La interrupción del proceso de aprendizaje y la falta de preparación académica personal sitúan a los individuos en una posición desfavorable en el mercado laboral, lo que inevitablemente impacta en sus ingresos económicos. En síntesis, aquellos con una educación más sólida tienen mayores oportunidades de empleo y reciben salarios más elevados que aquellos que carecen de formación o han interrumpido su educación.

Una consecuencia significativa de la exclusión del entorno académico es que la persona comienza a interactuar con un nuevo grupo social. Si este entorno social es poco saludable o fomenta valores y comportamientos irresponsables, los jóvenes se enfrentan a varios riesgos característicos de una sociedad corrupta, como la que estamos experimentando en la actualidad.

En otras palabras, esto implica que los individuos comienzan a desenvolverse en un entorno permeado por numerosas tentaciones y la posibilidad de verse envueltos en actividades como el consumo de drogas, la infracción de la ley y una dirección desorientada en la vida. Sin embargo, es crucial resaltar que, aunque al ingresar a la universidad los jóvenes pueden enfrentarse a problemas relacionados con el consumo de drogas o actividades ilegales, esto es válido siempre y cuando los estudiantes mantengan un compromiso serio con su educación y un deseo de superación personal. De este modo, la educación universitaria puede funcionar como una defensa contra la adicción a las drogas, el delito y la explotación sexual. (Aguirre Nativi y otros, 2011).

La deserción escolar da lugar a consecuencias sociales significativas cuya magnitud es difícil de cuantificar. Entre las causas subyacentes se encuentra una fuerza laboral con habilidades insuficientes y una falta de cualificaciones, lo que impide que las personas accedan a programas de capacitación gubernamentales, estatales o corporativos. El analfabetismo representa la manifestación más extrema de esta problemática. Asimismo, la disminución de la productividad laboral y su efecto en el crecimiento

económico restringido se identifican como costos sociales resultantes de la insuficiente educación, lo que podría contribuir a la deserción universitaria. Los gastos adicionales requeridos para financiar programas de asistencia social a sectores que carecen de capacidad para generar recursos por sí mismos también representan un costo social considerable.

Además, entre los costos vinculados con la deserción se encuentra la perpetuación intergeneracional de las desigualdades económicas y la pobreza, junto con su impacto adverso en la cohesión social. Esto dificulta los esfuerzos para fortalecer y consolidar la democracia.

En el ámbito de los costos individuales, es factible realizar cálculos considerando la proyección de los ingresos laborales menos favorables que los individuos adquirirán en el mercado de trabajo después de un periodo determinado de formación, contrastándolos con un estándar previamente establecido. En esencia, el costo individual alude a la suma de ingresos laborales que los jóvenes que abandonan prematuramente la educación universitaria no ganan a lo largo de su trayectoria profesional.

En resumen, los recursos destinados a este fin producirán beneficios tanto a nivel social como individual. Acceder a empleos de mayor calidad que requieran niveles más altos de educación también se traduce en menores tasas de desempleo, especialmente para aquellos que completan la universidad y optan por seguir avanzando en su formación. Además, puede resultar en una disminución de los salarios iniciales al buscar nuevas oportunidades laborales. (Aguirre Nativi y otros, 2011).

#### ***2.1.4.1. Estrategias para la prevención de la deserción***

Según lo señalado por González (2005), tal como se cita en el trabajo de Dzay Chulim & Narváz Trejo (2012), se distinguen dos categorías de estrategias para enfrentar el problema de la deserción: las generales y las específicas.

##### **Las estrategias comunes incluyen:**

Se propone impulsar una cultura de planificación en las instituciones universitarias, dado que actualmente carecen de una estrategia que considere plazos a corto, mediano y largo plazo. Se sugiere fomentar el desarrollo de programas técnicos o profesionales como alternativas educativas, simplificar los procedimientos para obtener un título, establecer un sistema de evaluación de las habilidades preuniversitarias necesarias en la escuela secundaria y promover la mejora continua tanto en los aspectos pedagógicos como administrativos de la institución. Entre las estrategias específicas se encuentran la revisión constante del plan de estudios institucional, así como la actualización y mejora continua del cuerpo docente.

##### **Dimensiones de la deserción universitaria**

La deserción universitaria se puede entender desde cuatro perspectivas principales: factores académicos, factores institucionales, factores socioeconómicos y factores personales.

##### **Dimensión 1: Factores académicos.**

Álvarez (1997) argumenta que los desafíos cognitivos, tales como el bajo desempeño académico, la retención, la carencia de disciplina y los métodos de aprendizaje deficientes, así como las deficiencias en el ámbito académico, como las dificultades en comprender los programas vinculados a la educación tradicional, la

insatisfacción en el proceso de aprendizaje debido a la ausencia de un entorno pedagógico apropiado, la falta de orientación profesional que se refleja en elecciones erróneas de carrera o institución educativa, y la ausencia de aptitud académica, son aspectos relevantes a considerar.

Según Carvajal & Rojas (2013), los factores académicos comprenden aspectos como las normativas institucionales, los horarios, las actividades, los exámenes, los talleres y las modalidades de evaluación.

**Ambiente social.** Aspectos como las características individuales de los profesores, estudiantes y personal administrativo son considerados, incluyendo la familiaridad, el compañerismo, la sociabilidad y la integración.

**Recursos de apoyo.** Se abordan las condiciones necesarias para llevar a cabo actividades académicas, como el acceso a bibliotecas, laboratorios, equipos audiovisuales y la organización de salidas de campo.

**Programas de bienestar estudiantil.** Incluyen una diversidad de servicios diseñados para asegurar el bienestar integral del estudiante, como atención psicológica, actividades recreativas y artísticas, así como asistencia con tutorías y asesoramiento.

**Ambiente académico.** Se hace referencia a la realización de diversas actividades como congresos, charlas, seminarios y eventos culturales.

## **Dimensión 2: Factores personales.**

Según Álvarez (1997), quien hace referencia a Baquero (1993), la deserción estudiantil puede atribuirse a una serie de factores tanto psicológicos, como la motivación, las emociones, la falta de adaptación y el incumplimiento de expectativas, como sociológicos, que incluyen la influencia del entorno familiar y otros grupos sociales como

amigos, compañeros de clase y vecinos. Además, existen otros motivos no clasificados aún, como la edad, la salud y la muerte.

De acuerdo a (Carvajal & Rojas, 2013) los factores personales implican:

**Rasgos generales.** Edad, estado civil, género, origen territorial grupo étnico.

**Personalidad.** El concepto de autoimagen, las intenciones personales, las habilidades que engloban los métodos de estudio, las expectativas, las preferencias, la perseverancia hacia metas, las demandas y aspiraciones individuales, junto con la disciplina, el nivel de compromiso, el entusiasmo y la motivación para aprender, son elementos fundamentales para considerar. Asimismo, se resalta la importancia de la personalidad en términos de habilidades para relacionarse, integrarse con otros y adaptarse a las normas sociales e institucionales. Las percepciones acerca de la vida social, la experiencia universitaria, la futura carrera profesional y los objetivos de vida también son aspectos significativos a tener en cuenta.

**Historia personal.** Como individuo integrado en la sociedad, su proceso de educación y la guía para la elección adecuada de carrera son aspectos fundamentales. La trayectoria académica del individuo desempeña un papel crucial al momento de embarcarse en la educación superior, reflejando su capital cultural y académico. Las deficiencias de conocimiento, dificultades en el aprendizaje, problemas con la comprensión de lectura y escritura, así como habilidades cognitivas limitadas y un desarrollo escaso del pensamiento lógico-matemático, junto con la carencia de hábitos de estudio, son desafíos frecuentes que enfrentan. Las decisiones cruciales que a menudo deben tomar están vinculadas con la orientación vocacional y profesional que recibieron al concluir su educación secundaria, la cual puede estar ausente en la vida del individuo.



**Actividades cotidianas.** Particularmente relacionado con la vida laboral, deportiva o artística del individuo, este puede dar prioridad a su carrera deportiva sobre sus compromisos académicos, influenciado por aspectos como la disponibilidad de tiempo, la dedicación, la perseverancia y la búsqueda de logros personales.

### **Dimensión 3: Factores socioeconómicos.**

Según Álvarez (1997), quien hace referencia a Baquero (1993), se señalan factores como los bajos ingresos familiares, la carencia de respaldo familiar, el desempleo y la incompatibilidad entre los horarios laborales y los de estudio como posibles factores socioeconómicos.

Para Carvajal & Rojas (2013) los factores socioeconómicos implican.

**Rasgos familiares.** Relacionados con el tamaño familiar, la trayectoria laboral de los miembros de la familia y los niveles educativos alcanzados por padres y hermanos.

**Capacidad económica.** Generalmente, los estudios sobre deserción universitaria tienden a analizar y enfocarse específicamente en este criterio. No obstante, consideramos que el aspecto económico, en términos de disponibilidad de recursos, recae principalmente en la familia, su capacidad financiera y la estabilidad laboral de sus miembros a lo largo del tiempo. La capacidad de acceder a estos recursos y comprometerse con ellos asegura la continuidad del individuo en el sistema educativo. La presencia o ausencia de estos recursos y la facilidad para obtenerlos son factores determinantes en situaciones de deserción, pero su adquisición está influenciada por la situación económica familiar o, en algunos casos, por las oportunidades laborales del individuo.

**Apoyo familiar.** Involucra garantizar la satisfacción de las necesidades fundamentales de un estudiante, como su alimentación, transporte, gastos cotidianos y

apoyo constante, además de proveer los recursos indispensables para llevar a cabo sus actividades académicas diarias.

**Valores de grupo familiar.** tales como el éxito, la solidaridad y el apoyo hacia los proyectos individuales, así como el nivel de independencia del individuo, son factores influyentes. Además, el ambiente familiar, que comprende los conflictos, su naturaleza, las relaciones emocionales entre los miembros y los eventos traumáticos como pérdidas, accidentes o enfermedades, también desempeñan un papel significativo.

#### **Dimensión 4: Factores institucionales**

De acuerdo con Álvarez (1997), quien hace referencia a Baquero (1993), se identifican factores institucionales como el cambio de centro educativo, problemas en la gestión administrativa, influencias negativas provenientes tanto de docentes como de otros miembros del personal institucional, así como la rigidez y obsolescencia de los programas académicos, y una calidad educativa deficiente, como posibles causas de la deserción estudiantil.

Según Carvajal & Rojas (2013), los elementos institucionales abarcan aspectos físicos vinculados con la disposición arquitectónica de la institución, tales como su estructura física, la calidad de las aulas, pizarras, pasillos, zonas recreativas y servicios sanitarios, así como las instalaciones de los comedores.

**Imagen institucional.** La imagen institucional se manifiesta en la percepción y reputación que tiene la institución educativa en diversos medios de comunicación como televisión, radio, prensa y páginas web. También, las referencias personales de amigos, vecinos y conocidos juegan un papel determinante en la valoración que el individuo hace de la institución.

**Condiciones financieras.** Comprenden tanto los gastos asociados a la carrera universitaria como las oportunidades de apoyo financiero para los estudiantes, que pueden incluir becas, opciones de financiamiento, posibilidades de obtención de créditos, descuentos y programas especiales para grupos como indígenas, personas afrodescendientes, personas con discapacidad y quienes destacan en el ámbito deportivo o académico.

**Calidad de los docentes.** Se relaciona con la formación académica de estos, sus características personales, los recursos utilizados en el proceso de enseñanza y métodos de enseñanza-aprendizaje.

**Pensum de carrera.** Que se refiere a los cursos ofrecidos en la carrera, incluyendo materias, seminarios opcionales, módulos, duración y organización del programa académico, así como la flexibilidad curricular y la distribución de los recursos a lo largo del tiempo.

**Modelo pedagógico.** Abarca tanto las estrategias implementadas por la institución como la manera en que el docente dirige el proceso educativo, según sus propios enfoques y sus métodos pedagógicos.

**Gestión administrativa.** Incluye aspectos como los procedimientos de pago, su calidad y eficiencia, así como la seguridad interna y la limpieza de las instalaciones).

## 2.2. Marco Conceptual

**Deserción:** La deserción se define como la interrupción definitiva y deliberada de la educación, pudiendo manifestarse en diferentes formas, como el abandono de un programa de estudios específico, la salida de una institución educativa o la renuncia al sistema educativo superior en su totalidad. (Mirez, 2014)

**Deserción universitaria:** Se trata del proceso en el que un estudiante deja su carrera profesional, ya sea por elección propia o por presión externa, debido a influencias tanto internas como externas que pueden tener un impacto positivo o negativo en el estudiante. (Acevedo, 2014)

**Factor institucional.** - El factor institucional es fundamental en la deserción universitaria, pues está relacionado con la calidad de la universidad, abarcando los espacios de estudio, la infraestructura y el nivel de los profesores. (Morí Sánchez, 2012)

**Factor individual.** - Según Mori (2012), El factor individual abarca una variedad de variables adicionales, como las características individuales, las metas personales, los deseos, las motivaciones y las preferencias vocacionales.

**Factor socioeconómico.** - El factor socioeconómico se caracteriza por su atención en elementos externos, abarcando lo social y lo económico. Aunque no se considera una causa principal de abandono, sí se reconoce como una circunstancia que puede generar desilusión y insatisfacción. (Morí Sánchez, 2012)

**Factores académicos:** Engloban desafíos cognitivos como el rendimiento académico deficiente, la repetición de cursos, la falta de disciplina y métodos de estudio inadecuados, así como falencias universitarias que incluyen dificultades en los programas académicos asociados con la enseñanza tradicional y la insatisfacción académica debido a la carencia de entornos pedagógicos idóneos para el aprendizaje. (Morí Sánchez, 2012)

## **2.3. Antecedentes Empíricos de la Investigación (estado del arte)**

### **2.3.1. Antecedentes Internacionales**

Montalvo Márquez, F., & Sánchez Pozo, N. (2023) en el estudio tuvo como objetivo “identificar los factores determinantes del abandono universitario mediante un

modelo logístico que permita predecir si un estudiante puede abandonar la universidad de manera prematura”. Se utilizó un modelo de regresión logística (LR) como técnica de análisis estadístico multivariante. Se examinó una muestra de 3821 registros de inscripciones en la UPEC y se determinó que el modelo propuesto tenía una precisión del 90%. Se identificó una relación inversa entre la deserción estudiantil y el tamaño de la familia, así como el nivel educativo de los padres, subrayando la necesidad de implementar medidas de apoyo específicas para estudiantes provenientes de hogares con menor nivel educativo y mayores responsabilidades familiares. Se concluyó que la deserción estudiantil en la UPEC es más frecuente durante el primer año de carrera, y que los factores más influyentes incluyen la edad, el tamaño de la familia, el nivel educativo y el rendimiento académico. Estos resultados ofrecen información valiosa para que las autoridades universitarias puedan implementar acciones preventivas orientadas a mejorar la retención y el éxito académico de los estudiantes.

Solís y otros (2022) en su estudio tuvo como objetivo principal “validar un modelo de regresión logística para la estimación de la deserción escolar en estudiantes de posgrado de la Universidad Técnica de Manabí, identificándose los principales factores influyentes que permitan tomar acciones sobre ellos”. Se recopilaron datos de variables predictoras disponibles en el Sistema de Gestión Académica, identificando a través de pruebas de Wald que el estado civil, la situación laboral, la edad y el salario eran altamente significativos. El modelo fue sometido a pruebas analíticas como el Chi cuadrado,  $R^2$  de Nagelkerke y Hosmer-Lemeshow, y se obtuvieron resultados significativos en todas ellas. Posteriormente, se realizó una clasificación de los resultados utilizando una muestra piloto, logrando un 98,6% de respuestas correctas. Finalmente, el modelo fue evaluado con una muestra de validación (30% de los datos), alcanzando un 99,06% de precisión.

Morocho Valarezo (2020) en su tesis de postgrado tuvo como objetivo principal “diseñar un modelo de Regresión Logística que permita la identificación de los factores que influyen en la deserción estudiantil de un Instituto Tecnológico del período 2015-2017 – Guayaquil”. Mediante un análisis descriptivo, se estableció que el motivo primordial de la deserción se relaciona con inconvenientes personales, seguido por la carencia de guía académica. Además, se utilizó la regresión logística para discernir los factores vinculados a la deserción estudiantil en el Instituto. De acuerdo con los resultados de este estudio, se evidenció que el principal factor influyente en la deserción es el estado civil de convivencia, lo cual incrementa las posibilidades de abandono en la Institución, con un coeficiente  $\beta$  de 2.34.

### **2.3.2. Antecedentes Nacionales**

Calero y otros (2023) en la investigación que tuvo como objetivo principal “determinar los factores que explican la deserción escolar en la región Huánuco, Perú, para el período 2010 – 2020”. Este estudio adoptó una aproximación aplicada, empleando un nivel explicativo y un diseño no experimental, fundamentado en dos marcos teóricos sobre la deserción escolar: el enfoque individual y el institucional. Se utilizaron datos de la Encuesta Nacional de Hogares para desarrollar un modelo de regresión logística. Los resultados revelaron que la probabilidad de deserción escolar aumenta significativamente en relación con diversos factores socioeconómicos, como pertenecer a un hogar de nivel socioeconómico bajo, residir en una familia monoparental y tener progenitores con antecedentes de abandono educativo básico. Además, aspectos personales como la edad, el género masculino y la necesidad de trabajar también se asociaron con una mayor probabilidad de deserción. En síntesis, se determinó que los principales factores socioeconómicos de la deserción escolar abarcan el nivel socioeconómico del hogar, los antecedentes de deserción de los padres y la estructura familiar parental, mientras que los

factores personales están relacionados con la edad, el género y la necesidad de trabajo del estudiante.

Álvarez Manrique (2021) en su tesis tiene como objetivo “determinar el factor que predomina en la deserción preuniversitario de una Universidad Privada de Ate, 2021”. Este estudio se fundamenta en un enfoque cuantitativo y sigue un diseño básico o puro con el propósito de examinar cómo los factores económicos, familiares, motivacionales y vocacionales influyen en la deserción de los estudiantes preuniversitarios. Se clasifica como no experimental, dado que no se manipularon las variables, y la recolección de datos se llevó a cabo en un solo momento con una muestra de 158 individuos. Se emplearon dos cuestionarios con escala de Likert, uno con 41 preguntas para evaluar los factores asociados a la deserción y otro con 12 preguntas para evaluar la deserción en sí. La contrastación de hipótesis se efectuó mediante regresión logística, puesto que la prueba de normalidad de los datos indicó que todas las variables y dimensiones mostraban una distribución no normal. Como resultado, se determinó que el factor económico poseía el mayor coeficiente ( $B = 0.104$ ), representando un 31.3% en la deserción de los estudiantes preuniversitarios en la Universidad Privada de Ate.

Gutiérrez (2022) en su investigación tuvo como objetivo principal “determinar mediante un modelo predictivo la deserción de estudiantes en el primer año de estudio en la Universidad Nacional Santiago Antúnez de Mayolo – 2022”. El análisis se llevó a cabo empleando una perspectiva cuantitativa y una metodología longitudinal con el objetivo de comprender los factores relacionados con la deserción estudiantil. Se aplicó un diseño pre experimental y se incluyeron todos los alumnos matriculados entre los semestres 2010-I y 2019-I, lo que totalizó 6440 casos. La muestra abarcó todos los casos disponibles para análisis predictivos. Los resultados señalaron que el modelo de Gradient Boosting fue el más eficiente, con una precisión del 94%, una sensibilidad del 86%, un puntaje F1

del 90%, una exactitud del 95%, un puntaje R-cuadrado del 75.12% en los datos de entrenamiento y un puntaje R2-cuadrado del 70.09% en los datos de prueba. En resumen, se determinó que la aplicación de algoritmos de machine learning posibilita predecir con alta exactitud la deserción estudiantil durante el primer año de estudio.

### **2.3.3. Antecedentes Locales**

Paucar y Quispe (2021) su investigación, tuvo como objetivo utilizar el análisis de supervivencia para explicar la deserción estudiantil y determinar los factores asociados con la deserción estudiantil, de los estudiantes de la escuela profesional de matemática, se trabajó con una población de 131 estudiantes que ingresaron en los semestres 2010 – 2012, y se les hizo un seguimiento de 10 semestres consecutivos. Y concluye que hubo una deserción de 61.8% del total de estudiantes; los factores asociados a la deserción de los estudiantes de la escuela profesional de matemática son: sexo, edad, ingreso económico familiar y ocupación del jefe de hogar.

Rondón (2020) en su estudio tuvo como objetivo “identificar de qué manera los factores asociados son determinantes en la deserción de los estudiantes en el instituto superior tecnológico privado Khipu Cusco 2016”. El análisis se basó en un enfoque descriptivo-correlacional con un diseño no experimental y un análisis cuantitativo. Los hallazgos indicaron una relación moderada y lineal entre los factores examinados y la deserción estudiantil, como lo demostró el coeficiente Tau B de Kendall, el cual obtuvo un valor de 0.605, tal como se detalla en la tabla 55. No obstante, se resaltó que estas conclusiones no pueden compararse directamente con investigaciones anteriores, debido a que la población estudiada no corresponde al ámbito universitario, lo que sugiere la singularidad del contexto analizado.



## CAPITULO III

### HIPÓTESIS Y VARIABLES

#### 3.1. Hipótesis

##### 3.1.1. Hipótesis General

Los factores influyentes en la deserción universitaria son los factores académicos e individuales obtenidas mediante la regresión logística binaria en los estudiantes de la escuela profesional de matemática de la UNSAAC, Cusco 2022-2023.

##### 3.1.2. Hipótesis Específicas

- Los indicadores académicos más influyentes en la deserción universitaria son la motivación y la complejidad de la matemática obtenidas mediante la regresión logística binaria en los estudiantes de la escuela profesional de matemática de la Universidad Nacional de San Antonio Abad del Cusco 2022-2023.
- Los indicadores institucionales más influyentes en la deserción universitaria son la orientación, seguimiento o asistencia y la comunicación afectiva obtenidas mediante la regresión logística binaria en los estudiantes de la escuela profesional de matemática de la Universidad Nacional de San Antonio Abad del Cusco 2022-2023.
- Los indicadores individuales más influyentes en la deserción universitaria son la satisfacción y la segunda opción obtenidas mediante la regresión logística binaria en los estudiantes de la escuela profesional de matemática de la Universidad Nacional de San Antonio Abad del Cusco 2022-2023.
- Los indicadores socioeconómicos más influyentes en la deserción universitaria es el trabajo del estudiante y el apoyo económico de la familia, obtenido mediante la

regresión logística binaria en los estudiantes de la escuela profesional de matemática de la Universidad Nacional de San Antonio Abad del Cusco 2022-2023.

### **3.2. Identificación de Variables e Indicadores**

#### **3.2.1. Variable Independiente**

Factores de la deserción universitaria

##### **Dimensiones**

- Factor académico
- Factor institucional
- Factor individual
- Factores socioeconómicos

#### **3.2.2. Variable dependiente**

Deserción universitaria

### 3.3. Operacionalización de variables

Variables	Definición conceptual	Definición operacional	Dimensiones	Indicadores	ITEMS	ESCALA
VD Deserción Universitaria	La deserción universitaria puede entenderse como abandono anticipado de un programa educativo antes de haber conseguido obtener el título o diploma, y se considera un período lo suficientemente prolongado como para eliminar la posibilidad de que el estudiante se reincorpore. (Soler et al. 2020, p.29).	Utiliza la Regresión Logística como un método estadístico para estimar la probabilidad de que un grupo de personas pertenezca a una categoría específica que sea relevante. En este enfoque, la variable dependiente Y se concibe como binaria, denotando la probabilidad de que cada observación pertenezca a una de dos clases posibles (0 o 1). La Regresión Logística establece un vínculo entre la variable objetivo y un conjunto de variables predictoras que ejercen influencia sobre ella, las cuales pueden adoptar tanto naturaleza cualitativa como cuantitativa (Ancco Cayllahua, 2017).	Los Estudiantes De la Escuela Profesional de Matemática	Si Deserto el Estudiante		
				No Deserto el Estudiante		
			Factor académico	Metodología de enseñanza	5	Ordinal Escala de tipo Likert  Nunca = 1
				Recursos didácticos	6	
				Motivación	7	
				Preparación universitaria Matemática compleja	8 9	

<p>VI</p> <p>Factores Asociados a la Deserción</p>	<p>Según Tinto (1997), la deserción abarca cuatro dimensiones o Factores.</p> <ol style="list-style-type: none"> <li>1. Factor Académico</li> <li>2. Factor Institucional</li> <li>3. Factor Individual</li> <li>4. Factor Socioeconómico</li> </ol>			Buen reconocimiento de la carrera	10	<p>Casi Nunca = 2</p> <p>A veces = 3</p> <p>Casi Siempre = 4</p> <p>Siempre = 5</p>	
				Factor institucional	Orientación, seguimiento o asistencia		11
					Relaciones positivas con docentes		12
					Ambiente o clima positivo		13
					Comunicación afectiva		14
					Nivel de seguridad		15
				Factor individual	Influencia de situación familiar		16
					Buenas relaciones familiares		17
					Satisfacción		18
					Segunda opción		19
				Factor socioeconómico	Buena situación financiera familiar		20
					Educación superior de padres		21
					Trabaja el estudiante		22
					Apoyo familiar		23
Beneficio o beca universitaria	24						

## CAPITULO IV

### METODOLOGÍA

#### **4.1. Ámbito de Estudio: Localización Política Y Geográfica**

##### **4.1.1. Área de Estudio**

La investigación tiene lugar en el departamento, provincia y distrito de Cusco, situados en la región sureste del país, con coordenadas geográficas de latitud -13.52264 y longitud -71.96734. (Ecotur Cusco, 2024)

##### **4.1.2. Lugar de Ejecución de la Investigación**

La investigación se llevó a cabo en la Escuela Profesional de Matemáticas de la Universidad Nacional Abad del Cusco durante el período comprendido entre 2022 - 2023. Esta institución está ubicada en la Avenida de la Cultura, número 733, en el distrito, provincia y departamento del Cusco, a una altitud de 3366 metros sobre el nivel del mar. Se trata de la cuarta institución universitaria más antigua de Perú y la quinta de América Latina.

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

El enfoque de esta investigación es cuantitativo, pues busca examinar y medir una relación específica entre dos variables en un grupo definido de personas.

El estudio se clasifica como investigación básica, siguiendo la definición de Carrasco (2005), lo cual implica que no busca una aplicación inmediata, sino más bien expandir y profundizar en el conocimiento científico existente sobre una práctica específica. Su objetivo principal es mejorar el contenido de las teorías científicas mediante un análisis minucioso. (p. 43).

Hernández y Baptista (2014) La investigación explicativa trasciende la mera descripción de conceptos o fenómenos, así como el simple establecimiento de relaciones entre ellos. Su propósito es indagar en las causas subyacentes de los hechos y fenómenos, tanto físicos como sociales. En esencia, busca proporcionar explicaciones sobre el porqué de un fenómeno y en qué circunstancias se manifiesta, así como sobre la relación entre dos o más variables.

El diseño de esta investigación sigue un enfoque transversal no experimental. Las investigaciones no empíricas son aquellas en las que no se manipulan deliberadamente las variables. En cambio, se observan los fenómenos tal como se presentan en la realidad con el propósito de analizarlos. (Hernández et al., 2014).

#### **4.3. Unidad de análisis**

La unidad de análisis son los estudiantes matriculados en la escuela de matemáticas de la UNSAAC Cusco durante los semestres 2022 I, 2022 II, 2023 I y 2023 II. Cada caso de estudio representa un conjunto de atributos específicos de un estudiante individual.

#### **4.4. Población de estudio**

Una población, o específicamente una población objetivo, se caracteriza como un grupo de elementos, sea de número finito o infinito, que comparten atributos similares y sobre los cuales se intenta extraer conclusiones en un estudio. La delimitación de la población objetivo está influenciada por la pregunta de investigación y los objetivos del estudio. (Arias, 2012)

Para este estudio el tamaño de la población se determinó de la siguiente manera, primero se solicitó al centro de cómputo la relación de aquellos estudiantes regulares, no regulares, ingresantes, egresantes y por ultimo los ponderados, por ejemplo para el

semestre 2022 – I se consideró a todos los estudiantes matriculados de ese semestre y a eso se sumó todos los nuevos ingresantes del semestre 2022 – II para luego restar aquellos estudiantes egresantes del semestre 2022 – I, y de la misma manera para los dos últimos semestres del 2023 I y II, obteniendo así una población de 244 casos de estudio que son los estudiantes de la Escuela Profesional de Matemática de la UNSAAC.

#### **4.5. Plan de Recolección de Datos**

El proceso de recolección de datos fue de la siguiente manera:

- Teniendo muy en clara la población objeto para luego, solicitar al centro de cómputo un reporte de los alumnos matriculados en los siguientes semestres, 2022 -I, 2022 -II, 2023-I y 2023 -II. En condición de Egresantes, Ingresantes, No Regulares, Ponderados y Regulares. Para poder encontrar el número total de estudiantes que fueron 244 y también pude observar aquellos estudiantes que ya no se matricularon en los semestres siguientes que son considerados estudiantes desertores que fueron 65.
- También solicite una lista de aquellos estudiantes en el mismo periodo que se matricularon a la escuela de matemática; pero que dejaron la carrera para seguir postulando e ingresar a otras escuelas profesionales como de Ing. Civil, Arquitectura, Informática, Electrónica y Geológica, etc.
- No tuve mayor inconveniente con los estudiantes que aún permanecen en la carrera por que los ubiqué en sus aulas de clase, donde tuve mayor dificultad fueron aquellos estudiantes desertores; pero a pesar de todo ello puede ubicarlos con la lista de reporte y la lista de estudiantes que ahora se encuentran en otras escuelas, y pude ubicarlos en sus nuevas escuelas

para aplicarles el cuestionario, y a los demás tuve que escribir a sus correos institucionales, contactarlos por llamadas telefónicas; pero siempre con el consentimiento de ellos y mantener en discreción sus datos personales.

- La aplicación del instrumento fue aplicada personalmente por el investigador a mediados y finales del semestre 2023 -II.

#### 4.6. Tamaño de Muestra

La muestra es un subconjunto finito y representativo que se selecciona de la población accesible. (Arias, 2012)

El tamaño de la muestra para este estudio es de 150 estudiantes de la escuela profesional de matemáticas de la Universidad Nacional de San Antonio Abad del Cusco.

$$n = \frac{Z_{(1-\alpha/2)}^2 * N * p(1-p)}{Z_{(1-\alpha/2)}^2 * p(1-p) + (N-1)(\varepsilon)^2}$$

#### Muestra probabilística

N=Tamaño poblacional N=244

Z=1,96 (Con 95% de confianza)

Error, e=5%=0,05

p=Probabilidad de éxito

$p = 50\% = 0.5$     $q = 50\% = 0.5$

$n = 149.5 \approx 150$ .



#### **4.7. Técnicas de selección de muestra**

La técnica que utilizamos para elegir la muestra en este estudio fue primero, el muestreo probabilístico aleatorio simple con una población  $N= 244$  y obteniendo una muestra de  $n= 150$ , para luego aplicar un Muestreo Probabilístico Estratificado, donde dividimos a la población en subpoblaciones o Estratos, los desertores y no desertores que son 65 y 179 respectivamente, así las muestras son halladas proporcionalmente al tamaño de cada Estrato, siendo la proporción 0.61475, entonces las muestras para cada estrato será  $n_D = 40$  y  $n_{\sim D} = 110$  donde:  $n_D$  es la muestra del número de estudiantes desertores y  $n_{\sim D}$  es la muestra del número de estudiantes no desertores respectivamente. siendo este tipo de muestreo que se ajusta mejor al tipo de datos que estamos analizando.

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

En esta investigación se emplearon dos técnicas principales: el análisis documental y la encuesta. Para recopilar datos, se accedió al centro de cómputo de la universidad. La encuesta se aplicó con el propósito de abordar todas las áreas relevantes para los estudiantes, generando conjuntos de datos completos para el análisis. Finalmente, se procesaron y analizaron los datos obtenidos mediante un conjunto final validado estadísticamente.

## CAPITULO V

### RESULTADOS

Mediante el software estadístico R y Rstudio versión 4.2.1 se han obtenido los resultados que se muestran en las siguientes tablas, con la finalidad de determinar los factores que influyen en la deserción universitaria de los estudiantes de la Escuela Profesional de Matemática de la UNSAAC, Cusco 2022-2023 mediante la regresión logística binaria. Para ello, se ha analizado la base de datos de estudiantes utilizando un instrumento previamente validado por juicio de expertos además se obtuvo un Alpha de Cronbach de 0.719 el cual indica que según Ruiz tiene una confiabilidad alta ya que se encuentra en un rango 0.61 – 0.80.

#### Descripción de la Muestra

**Tabla 1**

*Descripción de la muestra por edad y sexo*

		f	%
Sexo	Masculino	105	70.0%
	Femenino	45	30.0%
Edad	16 años	22	14.7%
	17 años	50	33.3%
	18 años	22	14.7%
	19 años	29	19.3%
	20 años	16	10.7%
	21 años	4	2.7%
	22 años	3	2.0%
	24 años	1	0.7%
	25 años	1	0.7%
	30 años	2	1.3%
	Mínimo	16 años	
	Máximo	30 años	
Promedio	18.19 años		

El análisis demográfico de los estudiantes de la Escuela Profesional de Matemática de la UNSAAC revela una marcada predominancia masculina (70.0%) sobre

la femenina (30.0%). La distribución etaria muestra una concentración significativa entre los 16 y 19 años, abarcando el 82% (123 estudiantes) de la muestra. La edad más frecuente es 17 años (50 estudiantes, 33.3%), seguida por 19 años (29 estudiantes, 19.3%) y 18 años (22 estudiantes, 14.7%). El promedio de edad se sitúa en 18.19 años, con un rango de 16 a 30 años, indicando la presencia de algunos estudiantes mayores. Esta estructura demográfica sugiere una población estudiantil predominantemente joven y masculina, con una variabilidad etaria que podría influir en los patrones de deserción universitaria. La concentración en edades tempranas y la disparidad de género son factores relevantes a considerar en el análisis de la deserción mediante el modelo de regresión logística propuesto en el estudio.

**Tabla 2**

*Valoración de las Dimensiones*

		F	%
Deserción	No desertó	110	73,3%
	Si desertó	40	26,7%
Factor académico	Bajo	49	32,7%
	Medio	66	44,0%
	Alto	35	23,3%
Factor institucional	Bajo	45	30,0%
	Medio	74	49,3%
	Alto	31	20,7%
Factor individual	Bajo	53	35,3%
	Medio	65	43,3%
	Alto	32	21,3%
Factor socioeconómico	Bajo	54	36,0%
	Medio	62	41,3%
	Alto	34	22,7%

En la tabla se observa que el 26,7% de los estudiantes desertaron, mientras que el 73,3% continuaron. Los factores académicos se distribuyen con un 32,7% con un nivel de valoración bajo, 44,0% de valoración medio y 23,3% de valoración alto. En cuanto a los factores institucionales, el 30,0% se encuentra en nivel de valoración bajo, el 49,3%

una valoración medio y el 20,7% en una valoración alto. Los factores individuales presentan un 35.3% nivel de valoración bajo, 43.3% un nivel de valoración medio y 21,3% un nivel de valoración alto. Finalmente, los factores socioeconómicos muestran un 36.0% nivel de valoración bajo, 41.3% nivel de valoración medio y un 22.7% nivel de valoración alto. Estos resultados indican que múltiples dimensiones afectan la deserción estudiantil, sugiriendo la necesidad de estrategias integrales que aborden estos factores para mejorar la retención universitaria.

### Estimación del modelo de regresión logística binaria

Se aplicó un modelo de regresión logística binaria para identificar los factores asociados a la deserción universitaria. Inicialmente, se incluyeron todos los factores en el modelo:

**Tabla 3**

*Estimación del Modelo*

```
Call:
glm(formula = Deserción ~ Factor_academi + Factor_instit + Factor_indiv +
     Factor_socioec, family = binomial, data = deser)

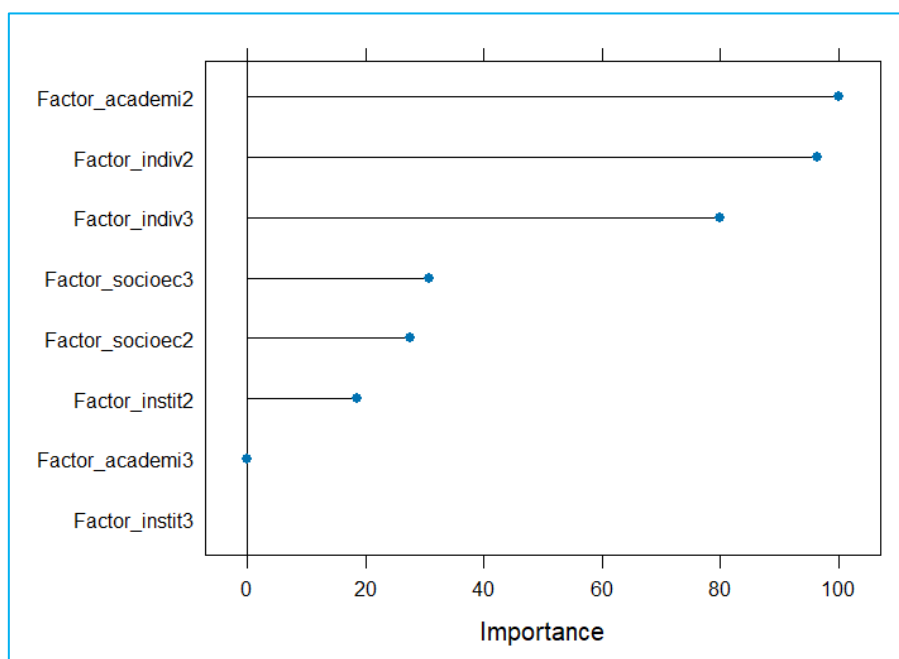
Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      1.877      0.677    2.77 0.00555 **
Factor_academi2  -2.589      0.659   -3.93 0.000086 ***
Factor_academi3 -20.478    2370.473  -0.01 0.99311
Factor_instit2    0.503      0.681    0.74 0.46035
Factor_instit3   -19.338    2607.328  -0.01 0.99408
Factor_indiv2    -2.940      0.777   -3.79 0.00015 ***
Factor_indiv3    -3.807      1.211   -3.14 0.00167 **
Factor_socioec2   0.688      0.632    1.09 0.27670
Factor_socioec3  -1.098      0.903   -1.22 0.22408
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### Interpretación:

- Este coeficiente es menor a uno (-2.589) y altamente significativo ( $p=0.000086$ ) indica que el nivel de valoración medio del factor académico reduce la probabilidad de deserción en 2.589, en comparación con el nivel de valoración bajo que dieron los estudiantes con respecto a los aspectos académicos.
- Este coeficiente es menor a uno (-2.940) y altamente significativo ( $p=0.00015$ ) indica que el nivel de valoración medio del factor individual reduce significativamente la probabilidad de deserción en 2.940, en comparación con la valoración bajo que dieron los estudiantes con respecto a los aspectos individuales.
- Este coeficiente es menor a uno (-3.807) y significativo ( $p=0.00167$ ) indica que el nivel de valoración alto del factor individual reduce la probabilidad de deserción en 3.807, en comparación con el nivel de valoración bajo que dieron los estudiantes con respecto a los aspectos individuales.

### Grafica 1

#### *Estimación del modelo*



Según los resultados algunos factores, como el Factor\_academi2, Factor\_indiv2 y Factor\_indiv3, tienen efectos significativos en la probabilidad de deserción, mientras que otros factores no muestran significancia estadística en este modelo.

### Selección de Variables

Se utilizó el método basado en el criterio de información de Akaike (AIC) para la selección de variables, resultando en un modelo que incluye los siguientes factores:

**Tabla 4**

*Selección de Variables*

```
Step: AIC=94.33
Deserción ~ Factor_academi + Factor_instit + Factor_indiv
```

	Df	Deviance	AIC
<none>		80.328	94.328
- Factor_instit	2	87.699	97.699
- Factor_indiv	2	106.378	116.378
- Factor_academi	2	112.741	122.741

**Tabla 5**

*Selección de Variables*

```
Call:
glm(formula = Deserción ~ Factor_academi + Factor_instit + Factor_indiv,
     family = binomial, data = deser)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	1.7315	0.5419	3.195	0.001396	**
Factor_academi2	-2.3695	0.6100	-3.884	0.000103	***
Factor_academi3	-20.2006	2445.3187	-0.008	0.993409	
Factor_instit2	0.6434	0.6718	0.958	0.338178	
Factor_instit3	-18.6050	2619.5688	-0.007	0.994333	
Factor_indiv2	-2.7651	0.7421	-3.726	0.000195	***
Factor_indiv3	-3.6201	1.1734	-3.085	0.002035	**

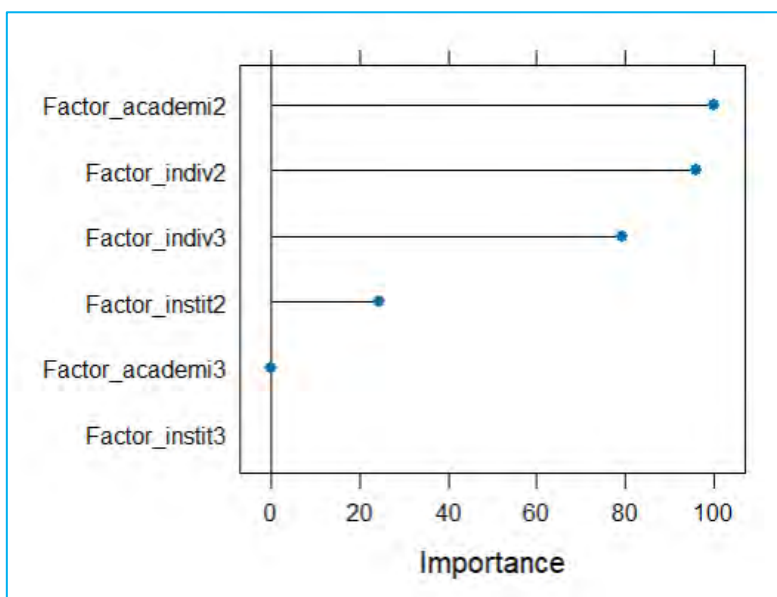
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Según los resultados de la interacción con AIC=94.33 el modelo de Regresión logística queda con las siguientes variables Factor\_academi2: (p=0.00010), significativo, Factor\_indiv2: (p=0.00019), significativo, Factor\_indiv3: (p=0.00203), significativo.

Este proceso de selección eliminó el factor socioeconómico del modelo final, indicando que no contribuía significativamente a la explicación de la deserción en presencia de los otros factores, como se muestra en la figura.

## Grafica 2

### Selección de Variable



## Tabla 6

### Odd Ratio

```
> ## tasa de ventajas e IC 95%
> exp(cbind(OR = coef(modelo_logistic), confint(modelo_logistic)))
waiting for profiling to be done...
              OR   2.5 %  97.5 %
(Intercept)  6.5e+00 1.93691 2.9e+01
Factor_academi2 7.5e-02 0.01785 2.5e-01
Factor_academi3 1.3e-09 0.00000 3.8e+38
Factor_instit2  1.7e+00 0.45418 6.8e+00
Factor_instit3  4.0e-09 0.00000 5.7e+48
Factor_indiv2  5.3e-02 0.00980 2.1e-01
Factor_indiv3  2.2e-02 0.00099 1.6e-01
```

El nivel Académico promedio con Odds Ratio de 0.075 indica que, tener un nivel de valoración académico medio presenta la probabilidad de que se reduzca la deserción universitaria, en comparación con tener un nivel de valoración académico bajo.

El nivel Individual medio con Odds Ratio de 0.053 indica que tener un nivel de valoración individual medio, presenta la probabilidad de que se reduzca la deserción universitaria en comparación con tener un nivel de valoración individual bajo.

El nivel Individual alto con Odds Ratio de 0.022 indica que tener un nivel de valoración individual alto, presenta la probabilidad de que se reduzca la deserción universitaria en comparación con tener un nivel de valoración individual bajo.

### Análisis del Factor Académico

**Tabla 7**

*Análisis del Factor Académico*

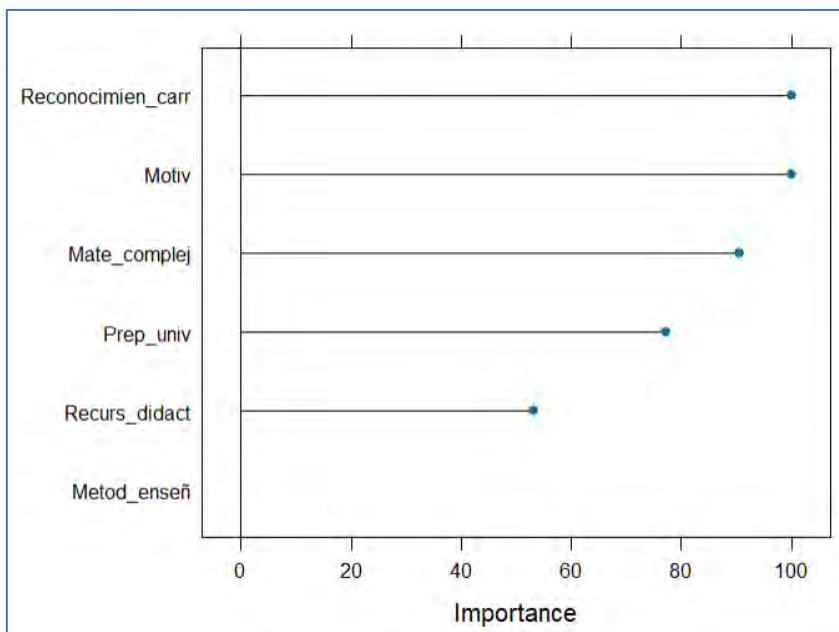
```
Call:
glm(formula = Deserción ~ Metod_enseñ + Recurs_didact + Motiv +
     Prep_univ + Mate_complej + Reconocimien_carr, family = binomial,
     data = deser)

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      7.79582    1.57778   4.941 0.000000777 ***
Metod_enseñ       0.04406    0.36620   0.120  0.90423
Recurs_didact    -0.61043    0.40627  -1.503  0.13296
Motiv            -0.75973    0.27907  -2.722  0.00648 **
Prep_univ        -0.43366    0.20398  -2.126  0.03350 *
Mate_complej     -0.68870    0.27790  -2.478  0.01320 *
Reconocimien_carr -0.88826    0.32628  -2.722  0.00648 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



### Grafica 3

#### Análisis del Factor Académico



Según los resultados del modelo de regresión logística, teniendo como variable dependiente a la variable deserción y al factor académico como variable regresora, se observa que los indicadores que más influyen en la deserción es el reconocimiento de la carrera, la motivación a permanecer, la complejidad de la matemática y la preparación universitaria, dichos indicador son influyente al 5% de significancia.

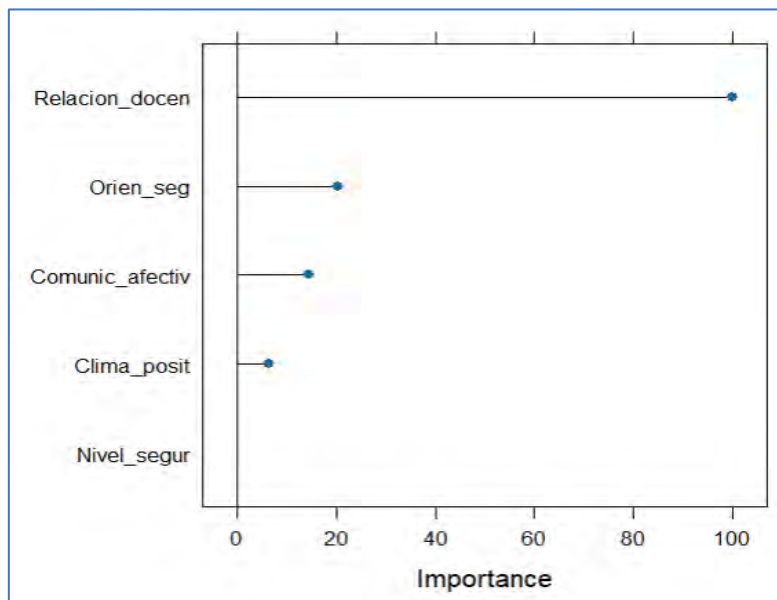
### Tabla 8

#### Análisis del Factor Institucional

```
Call:
glm(formula = Deserción ~ Orien_seg + Relacion_docen + Clima_posit +
     Comunic_afectiv + Nivel_segur, family = binomial, data = deser)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    2.9014    1.1673   2.486  0.01294 *
Orien_seg     -0.3928    0.3278  -1.198  0.23081
Relacion_docen -0.8447    0.2898  -2.914  0.00356 **
Clima_posit    0.2842    0.3156   0.901  0.36784
Comunic_afectiv -0.2404    0.2247  -1.070  0.28460
Nivel_segur   -0.1902    0.2495  -0.762  0.44577
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Grafica 4



Según los resultados del modelo de regresión logística, teniendo como variable dependiente a la variable deserción y al factor institucional como variable regresora, se observa que el indicador que influye en la deserción es la relación negativa con el docente, dicho indicador es influyente al 5% de significancia.

Tabla 9

*Análisis del Factor Individual*

## Coefficients:

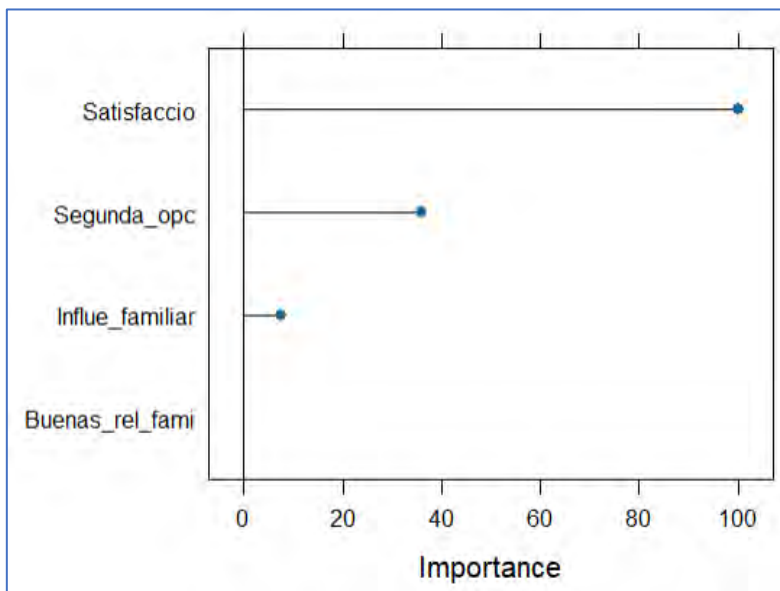
	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	2.85243	1.23624	2.307	0.0210 *
Influe_familiar	-0.14853	0.27361	-0.543	0.5872
Buenas_rel_fami	-0.04972	0.26037	-0.191	0.8486
Satisfaccio	-1.00653	0.21045	-4.783	0.00000173 ***
Segunda_opc	-0.60220	0.32605	-1.847	0.0648 .

---

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

(Dispersion parameter for binomial family taken to be 1)

Grafica 5



Según los resultados del modelo de regresión logística, teniendo como variable dependiente a la variable deserción y al factor individual como variable regresora, se observa que el indicador que influye en la deserción es la no satisfacción de la carrera, dicho indicador es influyente al 5% de significancia ( $p=0.00000173$ ).

Tabla 10

## Análisis del Factor Socioeconómico

```
Call:
glm(formula = Deserción ~ Buena_sit_fin + Educ_padres + Trabaja_estudi +
     Apoyo_familiar + Beneficio_beca, family = binomial, data = deser)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.03437    0.83738   1.235  0.2167
Buena_sit_fin -0.35483    0.23830  -1.489  0.1365
Educ_padres    0.19894    0.17261   1.153  0.2491
Trabaja_estudi 0.08228    0.17817   0.462  0.6442
Apoyo_familiar -0.35144    0.21180  -1.659  0.0971 .
Beneficio_beca -0.30015    0.17425  -1.723  0.0850 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Según los resultados del modelo de regresión logística, teniendo como variable dependiente a la variable deserción y al factor socioeconómico como variable regresora, se observa que no hay como tal ningún indicador que influya, pero los indicadores que

por ahí se acercan es el beneficio o beca y la buena situación financiera, dicho indicadores serian influyentes al 10% de significancia.

### **Calidad del ajuste del modelo**

Se realizó la prueba de Hosmer y Lemeshow para evaluar la bondad de ajuste del modelo. El resultado ( $p\text{-valor} > 0.05$ ) indica que no hay evidencia para rechazar la hipótesis nula, sugiriendo que el modelo se ajusta adecuadamente a los datos.

Según los resultados de la prueba Hosmer y Lemeshow el  $p\text{-valor} = 0.8001$  no se rechaza la hipótesis nula, por consiguiente, no hay diferencia entre observado y lo esperado. Por consiguiente, el modelo se ajusta a los datos.

### **Evaluación de la predicción**

Este paso es fundamental en el análisis de regresión logística. La función predict aplica el modelo ajustado a los datos para estimar la probabilidad de deserción de cada estudiante. Los valores obtenidos oscilan entre 0 y 1, donde:

Se utilizó un punto de corte de 0.5 para clasificar las predicciones:

Esta etapa transforma las probabilidades continuas en predicciones binarias de deserción. El umbral de 0.5 es un punto de partida estándar en regresión logística, basado en la idea de que, si la probabilidad es mayor al 50%, es más probable que el evento ocurra a que no ocurra.

**Tabla 11***Matriz de Confusión*

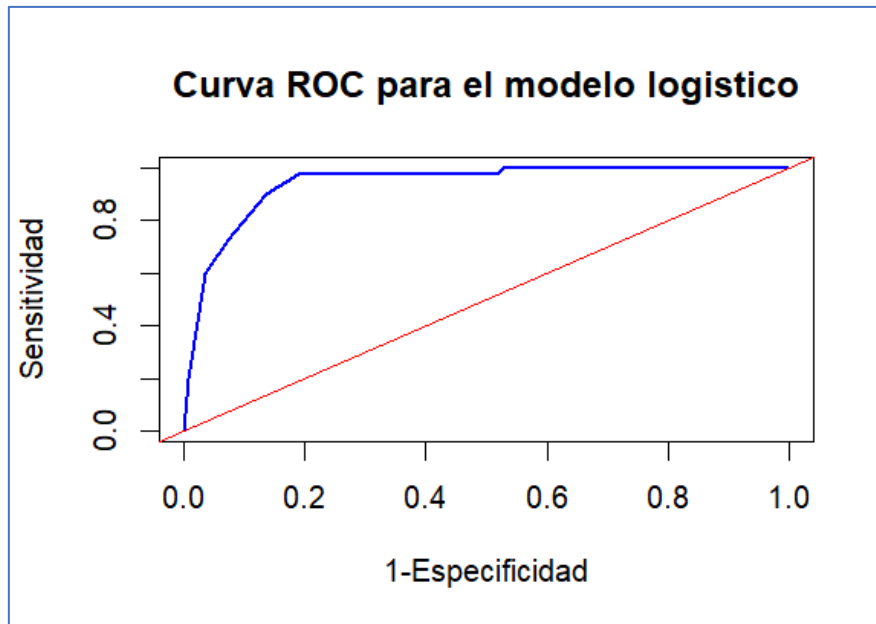
Confusion Matrix and Statistics		
	Reference	
Prediction	0	1
0	102	11
1	8	29
Accuracy : 0.8733		
95% CI : (0.8093, 0.922)		
No Information Rate : 0.7333		
P-Value [Acc > NIR] : 0.00002516		
Kappa : 0.6682		
McNemar's Test P-Value : 0.6464		
Sensitivity : 0.7250		
Specificity : 0.9273		
Pos Pred Value : 0.7838		
Neg Pred Value : 0.9027		
Prevalence : 0.2667		
Detection Rate : 0.1933		
Detection Prevalence : 0.2467		
Balanced Accuracy : 0.8261		
'Positive' Class : 1		

La matriz de confusión proporciona una visión detallada del rendimiento del modelo. Con una precisión del 87.33%, el modelo muestra un buen desempeño general. La sensibilidad de 72.50% indica que el modelo identifica correctamente más de tres cuartas partes de los casos reales de deserción, mientras que la especificidad de 92.73% sugiere una excelente capacidad para identificar casos de no deserción. El estadístico Kappa de 0.6682 indica un acuerdo sustancial entre las predicciones y los valores reales.

## Curva ROC y Área Bajo la Curva (AUC)

### Grafica 6

*Curva ROC o Aria Bajo la Curva*



La curva ROC y el área bajo la curva (AUC) son herramientas fundamentales para evaluar el rendimiento del clasificador. El AUC de 0.94 indica un excelente poder discriminativo del modelo, muy superior al 0.5 de un clasificador aleatorio. La curva ROC, que se sitúa considerablemente por encima de la línea diagonal (representativa de un clasificador aleatorio), confirma visualmente el buen rendimiento del modelo. Estos resultados sugieren que el modelo logístico desarrollado es altamente eficaz en distinguir entre los estudiantes que probablemente desertarán y los que no, basándose en los factores incluidos en el análisis.

### **Optimización del Punto de Corte**

El Índice de Youden se utiliza para optimizar el punto de corte del modelo, buscando maximizar tanto la sensibilidad como la especificidad. El punto de corte óptimo resultante de 0.3040 es menor que el valor predeterminado de 0.5, lo que refleja una decisión estratégica en el contexto de la deserción universitaria en la UNSAAC. Este nuevo umbral prioriza la identificación de más casos potenciales de deserción, incluso a costa de algunos falsos positivos. Esta elección está alineada con un enfoque preventivo, facilitando la detección temprana de estudiantes en riesgo para posibilitar intervenciones oportunas.

La reclasificación utilizando el nuevo umbral optimizado de 0.3040 refleja un enfoque más sensible en la detección de casos de deserción. Al reducir el punto de corte, más estudiantes serán clasificados como en riesgo de deserción. Este ajuste es coherente con una estrategia preventiva, donde se prefiere identificar un mayor número de casos potenciales de deserción para permitir intervenciones tempranas. La reclasificación modifica la distribución de las predicciones, potencialmente aumentando la capacidad del modelo para capturar casos de deserción que podrían haber sido pasados por alto con el umbral estándar.

**Tabla 12***Nueva Matriz de Confusión*

Confusion Matrix and Statistics		
	Reference	
Prediction	0	1
0	89	1
1	21	39
Accuracy : 0.8533		
95% CI : (0.7864, 0.9057)		
No Information Rate : 0.7333		
P-Value [Acc > NIR] : 0.0003233		
Kappa : 0.6765		
McNemar's Test P-Value : 0.00005104		
Sensitivity : 0.9750		
Specificity : 0.8091		
Pos Pred Value : 0.6500		
Neg Pred Value : 0.9889		
Prevalence : 0.2667		
Detection Rate : 0.2600		
Detection Prevalence : 0.4000		
Balanced Accuracy : 0.8920		
'Positive' Class : 1		

La nueva matriz de confusión refleja los cambios derivados del punto de corte optimizado. La sensibilidad aumenta significativamente a 97.50%, mejorando la detección de casos reales de deserción. Sin embargo, la especificidad disminuye a 80.91%, incrementando los falsos positivos. La precisión general se reduce ligeramente a 85.33%. El valor predictivo negativo aumenta a 98.89%, proporcionando mayor confianza en la clasificación de no desertores. Estos cambios indican una estrategia que prioriza la identificación amplia de estudiantes en riesgo, aceptando más falsos positivos para minimizar los casos de deserción no detectados. Estas métricas sugieren que el modelo es una herramienta valiosa para predecir la deserción universitaria en la UNSAAC.



## DISCUSIÓN DE RESULTADOS

El presente estudio ha logrado identificar, mediante un modelo de regresión logística binaria, los factores que influyen significativamente en la deserción universitaria de los estudiantes de la Escuela Profesional de Matemática de la UNSAAC. Los resultados revelan que los factores académicos e individuales son los más influyentes en la decisión de deserción, lo cual está en consonancia con los hallazgos de Montalvo Márquez y Sánchez Pozo (2023), quienes también identificaron el rendimiento académico como un factor crucial.

El modelo final, seleccionado mediante el criterio de información de Akaike (AIC), muestra una alta precisión (85.33%) en la predicción de la deserción, superando incluso el 83% de precisión reportado por Montalvo Márquez y Sánchez Pozo (2023). Esta alta precisión, respaldada por un área bajo la curva ROC (AUC) de 0.94, subraya la robustez del modelo y su potencial como herramienta predictiva, alineándose con la tendencia observada en estudios como el de Solís y otros (2022), que lograron un 98.6% de respuestas correctas en su modelo.

Es notable que, a diferencia de algunos estudios previos como el de Álvarez Manrique (2021), que identificó el factor económico como el más influyente, el presente estudio no encontró que el factor socioeconómico fuera significativo en el modelo final. Esto sugiere que, en el contexto específico de la Escuela de Matemática de la UNSAAC, los aspectos académicos e individuales tienen un peso mayor en la decisión de deserción que los factores económicos.

Los resultados indican que el reconocimiento de la carrera, la motivación y la preparación universitaria son los indicadores académicos más influyentes en la deserción. Esto coincide parcialmente con los hallazgos de Calero y otros (2023), quienes también

identificaron factores relacionados con el rendimiento académico como significativos. La importancia de estos indicadores sugiere que las intervenciones enfocadas en mejorar la orientación vocacional, fomentar la motivación y fortalecer la preparación académica podrían ser estrategias efectivas para reducir la deserción.

El análisis revela que la relación docente-estudiante es el principal indicador institucional que influye en la deserción. Este hallazgo resalta la importancia del entorno académico y la interacción entre profesores y alumnos, un aspecto que no fue prominente en los antecedentes revisados. Este resultado sugiere que las estrategias institucionales para mejorar la calidad de la relación docente-estudiante podrían tener un impacto significativo en la retención estudiantil.

La no satisfacción con la carrera emerge como el indicador individual más influyente en la deserción. Este hallazgo se alinea con los resultados de Morocho Valarezo (2020), quien identificó factores personales como determinantes en la deserción. La importancia de este indicador subraya la necesidad de implementar programas de orientación vocacional más efectivos y de proporcionar apoyo continuo a los estudiantes para ayudarles a alinear sus expectativas con la realidad de la carrera.

Aunque el factor socioeconómico no resultó significativo en el modelo final, el análisis individual de este factor reveló que el beneficio o beca y la situación financiera son indicadores influyentes al 10% de significancia. Esto contrasta con los hallazgos de Álvarez Manrique (2021), quien encontró que el factor económico era el más influyente. Esta diferencia podría atribuirse a las características específicas de la población estudiantil de la Escuela de Matemática de la UNSAAC o a la efectividad de los programas de apoyo financiero existentes.

## CONCLUSIONES

El modelo de regresión logística aplicado a la deserción universitaria en la Escuela Profesional de Matemática de la UNSAAC (2022-2023) determino que los factores académicos e individuales son significativos al 5%  $p=0.00086$  y  $p=0.00015$ , respectivamente.

Se determino que, del factor académico los indicadores más influyentes en la deserción son: El reconocimiento de la carrera, la motivación a permanecer en la carrera, la complejidad de las matemáticas y por último la preparación universitaria, son significativos 5% ( $p\text{-valor} < 0.05$ ). Los estudiantes con bajos niveles en estos factores tienen una mayor probabilidad de desertar. La distribución muestra que el 32.7% de los estudiantes se encuentra en un nivel de valoración bajo, el 44.0% en un nivel de valoración medio y el 23.3% en un nivel de valoración alto.

Se determino que, del factor institucional el indicador más influyente en la deserción es específicamente la relación docente-estudiante, es significativo al 5% ( $OR = 1.7, p < 0.05$ ). Los estudiantes con una mala relación docente-estudiante tienen casi el doble de probabilidad de desertar. La distribución de este factor es 30.0% en nivel bajo, 49.3% en nivel medio y 20.7% en nivel alto.

Se determino que, del factor individual, el indicador más influyente es la satisfacción con la carrera, es significativo al 5% ( $p < 0.05$ ). Los estudiantes con una baja satisfacción con su carrera tienen una mayor probabilidad de desertar. La distribución muestra que el 35.3% de los estudiantes se encuentra en un nivel bajo, el 43.3% en un nivel medio y el 21.3% en un nivel alto.

Aunque los indicadores socioeconómicos no fueron significativos en el modelo final ( $p > 0.05$ ), pero mostraron una tendencia de influencia al 10% de significancia. La

distribución de este factor es 36.0% en nivel bajo, 41.3% en nivel medio y 22.7% en nivel alto.

## RECOMENDACIONES

A la UNIVERSIDAD: Implementar un programa de Retención con un sistema integral de detección temprana de riesgo de deserción, con estrategias mediante el uso de la tecnología, actualmente el 100% de los estudiantes cuentan con un teléfono, y poder manejar ciertas aplicaciones como hacen algunas universidades extranjeras y sin ir muy lejos la Universidad Peruana de Ciencias Aplicadas (UPC), con el programa Dirección de Prospección Académica (DPA), y Tutoría Académica, haciendo un seguimiento y acompañamiento durante su etapa universitaria.

A la DIRECCIÓN ACADÉMICA: brindar talleres para estudiantes ingresantes darles a conocer sobre los beneficios que ofrece la universidad y la escuela de matemática y así mismo invitar a estudiantes egresados, que puedan compartir su experiencia en su paso por la universidad y como está les ha servido para su desarrollo personal, como también invitar a aquellos profesionales que vienen ejerciendo la carrera, en instituciones importantes como en la salud o entidades financieras. Para así mostrarles la importancia y el reconocimiento que tiene la Escuela de Matemática.

Para mejorar la calidad relación Docente-Estudiante y reducir la deserción, es crucial implementar programas de mentoría y formación docente, así como fomentar espacios de diálogo y retroalimentación.

Por último, Al CENTRO DE COMPUTO: manejar reportes con datos estadísticos más específicos y detallados de cada escuela, sobre el tema de deserción, y así obtener información para futuras investigaciones.

## BIBLIOGRAFÍA

- Acevedo, F. (2014). Una aproximación al fenómeno de desafiliación educativa en la transición hacia la educación superior en el interior de Uruguay. *Páginas de Educación*, 7(1), 180-210.
- Aguirre Nativi, P. A., Canjura Ramírez, M. H., & Serrano Alvarenga, N. H. (2011). *Factores sociopedagógicos que provocan la deserción de estudiantes en educación media en el Municipio de Soyapango*. Universidad Tecnológica de El Salvador. Obtenido de <http://biblioteca.utec.edu.sv:8080/jspui/handle/11298/137>
- Alvarez Manrique, L. (2021). *Factores relacionados a la deserción preuniversitario de una Universidad Privada de Ate, 2021*. Tesis de posgrado, Universidad Cesar Vallejo, Lima. Obtenido de [https://repositorio.ucv.edu.pe/bitstream/handle/20.500.12692/70249/Alvarez\\_MLD-SD.pdf?sequence=1&isAllowed=y](https://repositorio.ucv.edu.pe/bitstream/handle/20.500.12692/70249/Alvarez_MLD-SD.pdf?sequence=1&isAllowed=y)
- Álvarez, J. M. (1997). *Etiología de un sueño*. Universidad Autónoma de Colombia.
- Ancco Cayllahua, A. A. (2017). *Factores asociados al rendimiento académico en los cursos de Matemática Básica y Cálculo I de los alumnos ingresantes de la FCM-UNMSM utilizando regresión logística binaria*. Tesis de pre grado, Universidad Nacional Mayor de San Marcos, Lima. Obtenido de [https://cybertesis.unmsm.edu.pe/bitstream/handle/20.500.12672/14574/Ancco\\_ca.pdf?sequence=1&isAllowed=y](https://cybertesis.unmsm.edu.pe/bitstream/handle/20.500.12672/14574/Ancco_ca.pdf?sequence=1&isAllowed=y)
- Attinasi, L. C. (1986). *Getting in: Mexican american students perceptions of their college - going behavior with implications for their freshman year persistence in the university*. Obtenido de [www.law.uh.edu](http://www.law.uh.edu): <https://www.law.uh.edu/ihelg/monograph/87-4.pdf>
- Bean, J. P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in Higher Education*, 12(2), 155–187. Obtenido de <https://doi.org/10.1007/BF00976194>
- Behr, A., Giese, M., Tegum, H., & Theune, K. (2020). Early Prediction of University Dropouts-A Random Forest Approach. *Jahrbücher für Natl. und Statistik*, 743-789. doi:<https://doi.org/10.1515/jbnst-2020-0006>
- Braxton, J. M., Shaw Sullivan, A. V., & Johnson, R. M. (28 de Febrero de 1997). *Appraising Tint's Theory of Collage Student Departure*. Obtenido de [books.google.es](https://books.google.es): [https://books.google.es/books?hl=es&lr=&id=twYjXLd6pyUC&oi=fnd&pg=PA107&dq=Braxton+\(1997\)+&ots=D9hbS1X740&sig=gqNGfyrS99O8CWqgEUyPO8\\_pt3g#v=onepage&q&f=false](https://books.google.es/books?hl=es&lr=&id=twYjXLd6pyUC&oi=fnd&pg=PA107&dq=Braxton+(1997)+&ots=D9hbS1X740&sig=gqNGfyrS99O8CWqgEUyPO8_pt3g#v=onepage&q&f=false)
- Calero, R., Sosa-Cagna, M., Lino-Zuñiga, L., & Ponciano-Navarro, J. (2023). Factores determinantes de la deserción escolar en la región Huánuco, Perú. *Desafíos*, 14(2). doi:<https://doi.org/10.37711/desafios.2023.14.2.401>
- Calero, R., Sosa-Cagna, M., Lino-Zuñiga, L., & Ponciano-Navarro, J. (2023). *Factores determinantes de la deserción escolar en la región Huánuco, Perú*. Artículo, Universidad de Huánuco, Huánuco. doi:<https://doi.org/10.37711/desafios.2023.14.2.401>

- Carvajal Olaya, P., Trejos Carpintero, A. A., & Gómez Molina, R. H. (2009). *Estrategias de diagnóstico y*. Pereira: Publiprint. Obtenido de <https://www.worldcat.org/title/estrategias-de-diagnostico-y-acompañamiento-una-experiencia-exitosa-para-la-permanencia-estudiantil-en-le-universidad-tecnologica-de-pereira-en-la-direccion.correcta/oclc/916300988?referer=di&ht=edition>
- Carvajal, A., & Rojas, R. (2013). *Factores que inciden en la deserción de los estudiantes en la USTA Colombia*. Universidad Santo Thomas.
- Dávila, R., Agüero, E., Portillo, H., & Quimbita, O. (2022). Deserción universitaria de los estudiantes de una universidad peruana. *Revista Universidad y Sociedad*, 2(14), 421-427. Obtenido de <https://rus.ucf.edu.cu/index.php/rus/article/view/2722>
- Dzay Chulim, F., & Narváz Trejo, O. M. (2012). *La deserción escolar desde la perspectiva estudiantil*.
- Ecotur Cusco. (2024). *Estaciones y Clima de Cusco*. Obtenido de <https://ecotourcusco.com/es/estaciones-y-clima-de-cusco/>
- Girón Azurdia, R. F. (2014). *"Factores De Riesgo Que Ocasionaron La Deserción De Estudiantes De La Facultad De Ingeniería De La Universidad Rafael Landívar"*. Universidad Rafael Landívar. Obtenido de <http://biblio3.url.edu.gt/Tesario/2014/05/83/Giron-Freddy.pdf>
- González, L. (2005). *Estudio sobre la repetición y deserción en la educación superior Chilena*. Digital Observatory for higher education in Latin America and The Caribbean, IESALC-UNESCO.
- González, M., Miquelena, L., Rebolledo, A., & Vásquez, H. (2018). Deserción Estudiantil en la Universidad del Zulia ¿Problema Coyuntural? *Revista Venezolana de Gerencia, Edición Especial 1*(23), 403-431.
- Gutierrez Churata, H. (2022). *Modelo predictivo para la deserción de estudiantes en el primer año de estudio en la Universidad Nacional Santiago Antúnez de Mayolo, Huaraz – 2022*. Tesis de pre grado, Universidad Nacional Santiago Antúnez De Mayolo, Huaraz. Obtenido de [https://repositorio.unasam.edu.pe/bitstream/handle/UNASAM/5361/T033\\_71285563\\_T.pdf?sequence=1&isAllowed=y](https://repositorio.unasam.edu.pe/bitstream/handle/UNASAM/5361/T033_71285563_T.pdf?sequence=1&isAllowed=y)
- Hosmer, W., & Lemeshow, S. (2000). *Applied Logistic Regression. Second Edition. A Wiley-Interscience Publication*. John Wiley & Sons .
- Mairata, J. (2010). *Análisis del Perfil y de las Causas de Abandono en la Educación Superior*. Universidad de las Illes Balears.
- Mirez, A. (2014). *La deserción estudiantil en la Facultad de Ciencias Económicas y Empresariales de la Universidad de Piura y su impacto en los resultados económicos*. Universidad de Piura. Universidad de Piura. Obtenido de <https://pirhua.udep.edu.pe/handle/11042/1792>
- Montalvo Márquez, F. J., & Sánchez Pozo, N. N. (30 de Junio de 2023). *Modelo de predicción de deserción: Un estudio de caso de estudiantes de la Universidad Politécnica Estatal del Carchi*. Artículo, Universidad Politécnica Estatal del Carchi, Carchi - Ecuador. doi:<https://doi.org/10.55204/pcc.v3i2.e34>

- Morí Sánchez, N. (2012). *Deserci n universitaria en estudiantes de una Universidad Privada de Iquitos*. Obtenido de <https://revistas.upc.edu.pe/index.php/docencia/article/view/42>
- Morocho Valarezo, K. (2020). *Modelo de regresi n log stica de la deserci n estudiantil en un instituto tecnol gico en el cant n daule*. Tesis de posgrado, Escuela Superior Polit cnica Del Litoral, Guayaquil.
- Morocho Valarezo, K. (2020). *Modelo de regresi n log stica de la deserci n estudiantil en un instituto tecnol gico en el cant n daule*. Tesis de posgrado, Escuela Superior Polit cnica Del Litoral, Guayaquil. Obtenido de <https://www.dspace.espol.edu.ec/bitstream/123456789/51577/1/T-110157.pdf>
- Narv rez Barros, M. A., & Barrag n Reyes, G. E. (2015). *An lisis sobre la deserci n estudiantil en la universidad Polit cnica Salesiana, Sede Guayaquil: caso de las carreras de Administraci n de Empresas y Contabilidad y Auditor a periodo de aplicaci n 2007-2012*. Universidad de Ciencias Administrativas y Econ micas Sede Guayaquil. Obtenido de <http://dspace.ups.edu.ec/handle/123456789/9054>
- Rond n Tamayo, C. (2020). *Factores Asociados A La Desercion Estudiantil En El Instituto De Educacion Superior Tecnologico Khipu Cusco - 2016*. Tesis de pos grado, Universidad Nacional De San Antonio Abad Del Cusco , Cusco. Obtenido de [https://repositorio.unsaac.edu.pe/bitstream/handle/20.500.12918/5365/253T202010\\_09\\_TC.pdf?sequence=1&isAllowed=y](https://repositorio.unsaac.edu.pe/bitstream/handle/20.500.12918/5365/253T202010_09_TC.pdf?sequence=1&isAllowed=y)
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Mach. Intell. 1*, 206–215. Obtenido de <https://www.nature.com/articles/s42256-019-0048>
- Sol s Ventura, J., Quiroz Fern ndez, S., & Fosado T llez, O. (2022). *Modelo de regresi n log stica para la estimaci n de la deserci n escolar del posgrado en la universidad t cnica de manab , ecuador*. Articulo, Universidad T cnica de Manab . doi:<https://doi.org/10.33936/revbasdelaciencia.v7i3.5197>
- Tinto, V. (1975). Dropout from Higher Education: A Theoretical Synthesis of Recent Research. *Review of Educational Research*, 45(1), 89–125. Obtenido de <https://doi.org/10.3102/00346543045001089>
- UNESCO. (2006). Informe sobre la Educaci n Superior en Am rica Latina y el Caribe 2000-2005 Talca Repitencia y deserci n universitaria en Am rica Latina. *Organizaci n de la Naciones Unidas para la Educaci n la Ciencias y la Cultura /Instituto Internacional Para La Educaci n Superior en Am rica Latina y el Caribe*, 156-168.

# ANEXOS

UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO  
FACULTAD DE CIENCIAS

ESCUELA PROFESIONAL DE MATEMÁTICA

## CUESTIONARIO

**Estimado(a) Estudiante**

La presente encuesta es anónima forma parte de un proyecto de investigación, el cual tiene como objetivo determinar los factores que influyen en la deserción universitaria de los estudiantes de la Escuela Profesional de Matemática de la UNSAAC, con el modelo de Regresión Logística, Cusco 2022-2023. Por tal razón solicito diligenciar la siguiente encuesta con sinceridad y responsabilidad, ya que dicha información solo se utilizará para fines académicos.

### **I. Datos generales**

1. **Código del estudiante:** .....

2. **Género:**

a. Femenino

b. Masculino

3. **¿Cuál fue su edad al momento de su ingreso a la universidad?**

.....

### **II. Deserción Universitaria**

4. **¿Usted desertó de sus estudios de forma definitiva de la escuela profesional de matemática?**

1. Si

2. No



### III. Factores de deserción universitaria

Lea cada enunciado y posteriormente marque con una (x) en el recuadro correspondiente según su criterio se ajuste a la verdad. Nunca = 1, Casi Nunca = 2, A veces = 3, Casi siempre = 4, Siempre = 5						
Nº	Ítems	1	2	3	4	5
<b>Factor Académico</b>						
5.	Estas o estabas de acuerdo con la metodología de enseñanza de los docentes.					
6.	Estas o estabas de acuerdo con los recursos didáctica utilizados por los docentes.					
7.	Los docentes de tu carrera incentivan y motivan a los estudiantes a permanecer en la carrera.					
8.	Recibiste preparación universitaria antes de ingresar a la universidad.					
9.	Consideras que los cursos de matemáticas son complejos.					
10.	La escuela de matemáticas tiene buen reconocimiento frente a las otras carreras y un buen campo laboral.					
<b>Factor Institucional</b>						
11.	Evalúa la frecuencia con la que sientes que la universidad te proporciona orientación, seguimiento o asistencia durante tus estudios.					
12.	Por favor, indique con qué frecuencia experimenta relaciones positivas con los docentes de la universidad					
13.	Indique con qué frecuencia experimenta un ambiente o clima positivo en la universidad.					
14.	Evalúe la frecuencia con la que tiene una comunicación efectiva con tus compañeros de la universidad.					
15.	Indique con qué frecuencia percibe un alto nivel de seguridad en la universidad.					
<b>Factores Individuales</b>						
16.	Indique la frecuencia con la que su situación familiar (esposas, hijos u otros) influye en su vida diaria.					
17.	Indique con qué frecuencia experimentas buenas relaciones familiares en tu casa.					
18.	Evalúa la frecuencia de satisfacción con la carrera que estudia o estudiaba.					
19.	Indique con qué frecuencia eligió la segunda opción como modalidad de ingreso a su carrera profesional.					
<b>Factores Socioeconómicos</b>						
20.	Su familia goza de una buena situación financiera					
21.	Tus padres o las personas con las que convive cuentan con educación superior.					
22.	Realiza actividad económica para pagar sus estudios de carrera.					
23.	Tiene el apoyo de su familia para realizar sus estudios					
24.	Tienes algún tipo de beneficio o beca de la universidad u otra institución					

## VALIDACION DE INSTRUMENTOS MEDIANTE JUICIO DE EXPERTOS

**SOLICITUD:** Validación de instrumento de recojo de información.

Dr.: DIONICIO CARDENAS CANCHA

Yo Franklin Cuchillo Valverde, identificado con el DNI N° 74401449 alumno de la Escuela Profesional de Matemática, me dirijo a usted con el debido respeto y le manifiesto lo siguiente:

Que siendo requisito indispensable el recojo de datos necesarios para la tesis que viene elaborando titulada: **“Factores que influyen en la deserción universitaria de los estudiantes de la Escuela Profesional de Matemática de la UNSAAC, con el Modelo de Regresión Logística, Cusco 2022-2023”**, solicito a Ud. Se sirva validar el instrumento adjunto, bajo los criterios académicos correspondientes. Para este efecto adjunto los siguientes documentos:

- Ficha de evaluación
- Instrumento

Por tanto:

A usted, ruego acceder mi petición.

Cusco, 10 de noviembre del 2023

---

DNI: 74401449

## VALIDACIÓN DE INSTRUMENTOS

1.1. Apellidos y Nombres: DIONICIO CÁRDENAS CANCHA

1.2. Cargo e institución donde labora: PROFESOR PRINCIPAL DE LA UNSAAC

1.3. Especialidad o línea de investigación: DOCTOR EN CIENCIAS DE LA EDUCACIÓN

1.4. Nombre del instrumento motivo de evaluación: \_\_\_\_\_

1.5. Autor(a) de Instrumento: Elaboración propia

### I. ASPECTOS DE VALIDACIÓN

CRITERIOS	INDICADORES	INACEPTABLE						MÍNIMAMENTE ACEPTABLE			ACEPTABLE			
		40	45	50	55	60	65	70	75	80	85	90	95	100
1. CLARIDAD	Esta formulado con lenguaje comprensible.							X						
2. OBJETIVIDAD	Esta adecuado a las leyes y principios científicos.								X					
3. ACTUALIDAD	Esta adecuado a los objetivos y las necesidades reales de la investigación.							X						
4. ORGANIZACIÓN	Existe una organización lógica.								X					
5. SUFICIENCIA	Toma en cuenta los aspectos metodológicos esenciales								X					
6. INTENCIONALIDAD	Esta adecuado para valorar las variables de la Hipótesis.								X					
7. CONSISTENCIA	Se respalda en fundamentos técnicos y/o científicos.								X					
8. COHERENCIA	Existe coherencia entre los problemas objetivos, hipótesis, variables e indicadores.							X						
9. METODOLOGÍA	La estrategia responde una metodología y diseño aplicados para lograr probar las hipótesis.							X						
10. PERTINENCIA	El instrumento muestra la relación entre los componentes de la investigación y su adecuación al Método Científico.							X						

### II. OPINIÓN DE APLICABILIDAD

- El Instrumento cumple con los Requisitos para su aplicación
- El Instrumento no cumple con los requisitos para su aplicación



### IV. PROMEDIO DE VALORACIÓN

75

FIRMA DEL ESPECIALISTA  
DIONICIO CÁRDENAS CANCHA  
Nro. de colegiatura: 85

## VALIDACIÓN DE INSTRUMENTOS

- 1.1. Apellidos y Nombres: CLETO DE LA TORRE DUEÑAS  
 1.2. Cargo e institución donde labora: DOCENTE PRINCIPAL DE LA UNSAAC  
 1.3. Especialidad o línea de investigación: DOCTOR EN CIENCIAS  
 1.4. Nombre del instrumento motivo de evaluación: \_\_\_\_\_  
 1.5. Autor(a) de Instrumento: Elaboración propia

## I. ASPECTOS DE VALIDACIÓN

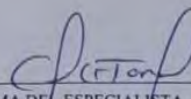
CRITERIOS	INDICADORES	INACEPTABLE						MÍNIMAMENTE ACEPTABLE			ACEPTABLE			
		40	45	50	55	60	65	70	75	80	85	90	95	100
1. CLARIDAD	Esta formulado con lenguaje comprensible.											X		
2. OBJETIVIDAD	Esta adecuado a las leyes y principios científicos.											X		
3. ACTUALIDAD	Esta adecuado a los objetivos y las necesidades reales de la investigación.												X	
4. ORGANIZACIÓN	Existe una organización lógica.										X			
5. SUFICIENCIA	Toma en cuenta los aspectos metodológicos esenciales											X		
6. INTENCIONALIDAD	Esta adecuado para valorar las variables de la Hipótesis.												X	
7. CONSISTENCIA	Se respalda en fundamentos técnicos y/o científicos.											X		
8. COHERENCIA	Existe coherencia entre los problemas objetivos, hipótesis, variables e indicadores.											X		
9. METODOLOGÍA	La estrategia responde una metodología y diseño aplicados para lograr probar las hipótesis.											X		
10. PERTINENCIA	El instrumento muestra la relación entre los componentes de la investigación y su adecuación al Método Científico.											X		

## II. OPINIÓN DE APLICABILIDAD

- El Instrumento cumple con los Requisitos para su aplicación
- El Instrumento no cumple con los requisitos para su aplicación



## IV. PROMEDIO DE VALORACIÓN

  
 FIRMA DEL ESPECIALISTA  
 Nombres y Apellidos  
 Nro. de colegiatura

## VALIDACIÓN DE INSTRUMENTOS

- 1.1. Apellidos y Nombres: APARICIO ARENAS Karla ZELMIRA.  
 1.2. Cargo e institución donde labora: DOCENTE - UNSAAC.  
 1.3. Especialidad o línea de investigación: ESTADÍSTICA,  
DESECCION.  
 1.4. Nombre del instrumento motivo de evaluación:  
 1.5. Autor(a) de Instrumento: Elaboración propia

## I. ASPECTOS DE VALIDACIÓN

CRITERIOS	INDICADORES	INACEPTABLE					MINIMAMENTE ACEPTABLE			ACEPTABLE				
		40	45	50	55	60	65	70	75	80	85	90	95	100
1. CLARIDAD	Esta formulado con lenguaje comprensible.											X		
2. OBJETIVIDAD	Esta adecuado a las leyes y principios científicos.												X	
3. ACTUALIDAD	Esta adecuado a los objetivos y las necesidades reales de la investigación.												X	
4. ORGANIZACIÓN	Existe una organización lógica.												X	
5. SUFICIENCIA	Toma en cuenta los aspectos metodológicos esenciales											X		
6. INTENCIONALIDAD	Esta adecuado para valorar las variables de la Hipótesis.												X	
7. CONSISTENCIA	Se respalda en fundamentos técnicos y/o científicos.												X	
8. COHERENCIA	Existe coherencia entre los problemas objetivos, hipótesis, variables e indicadores.												X	
9. METODOLOGÍA	La estrategia responde una metodología y diseño aplicados para lograr probar las hipótesis.											X		
10. PERTINENCIA	El instrumento muestra la relación entre los componentes de la investigación y su adecuación al Método Científico.												X	

## II. OPINIÓN DE APLICABILIDAD

- El Instrumento cumple con los Requisitos para su aplicación
- El Instrumento no cumple con los requisitos para su aplicación

## IV. PROMEDIO DE VALORACIÓN



94

UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO  
 FACULTAD DE CIENCIAS  
 Dra. Karla Zelmira Aparicio Arenas  
 COMAP-496

FIRMA DEL ESPECIALISTA  
 Nombres y Apellidos  
 Nro. de colegiatura

## FUNCIONES QUE SE UTILIZÓ EN R STUDIO

### Curva ROC Aria Bajo la Curva

```
> analysis ## area bajo la curva

call:
roc.default(response = deser$Deserción, predictor = yprob)

Data: yprob in 110 controls (deser$Deserción 0) < 40 cases (deser$Deserción 1).
Area under the curve: 0.94
> MLmetrics::AUC(y_pred = yprob,y_true =as.numeric(deser$Deserción)-1 ) # area bajo la curva
[1] 0.94
```

```
library(pROC)
analysis <- roc(response=deser$Deserción, predictor=yprob)
MLmetrics::AUC(y_pred = yprob,y_true =as.numeric(deser$Deserción)-1 )

plot(1-analysis$specificities,analysis$sensitivities,type="l",
     ylab="Sensitividad",xlab="1-Especificidad",col="blue",lwd=2,
     main = "Curva ROC para el modelo logistico")
abline(a=0,b=1, col = "red")
```

```
e <- cbind(analysis$thresholds,analysis$sensitivities+analysis$specificities-1)
opt_t <- subset(e,e[,2]==max(e[,2]))[,1]
yprob2<-predict(modelo_logistic,type="response")

ypred2 <- as.numeric(yprob2 >= 0.3040 )
ypred2 <- factor(ypred2, labels = levels(deser$Deserción))

library(caret)
confusionMatrix(data = ypred2,reference = deser$Deserción, positive = "1")
```

### Evaluación de la predicción

```
yprob<-predict(modelo_logistic,type="response")
head(yprob)
```

```
ypred <- as.numeric(yprob >= 0.5 )
ypred <- factor(ypred, labels = levels(deser$Deserción))
```

```
library(caret)
confusionMatrix(data = ypred,reference = deser$Deserción, positive = "1")
```

### Evaluación de la predicción

```
Hosmer and Lemeshow goodness of fit (GOF) test

data: modelo_logistic$y, fitted(modelo_logistic)
X-squared = 4.5926, df = 8, p-value = 0.8001
```

```

> yncrad <- as.numeric(yncrad >= 0.5 )
> yncrad
[1] 0 0 1 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 1 0 0 0 1 0 0 1 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
[90] 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 0 0 1 0 0

> yncob
      1          2          3          4
5      6      7      8      9     10     11
0.1314046037796580967960 0.2235379387656113714389 0.5013604745398737572160 0.0000000004855227579430
0.8496002197790717413994 0.000000000000002220446 0.000000000000002220446
12      13      14      15      16      17      18
0.2235379387656113714389 0.0262210111496524837793 0.0321984835864497082047 0.0595423933648906306138
0.000000001141583333245 0.0262210111496524837793 0.0000000095267770211087
19      20      21      22      23      24      25
0.5013604745398737572160 0.3457015569155127421297 0.0000000001176749447137 0.2623768994670613308173
0.0139524017064684060369 0.000000001141583333245 0.0321984835864497082047
26      27      28      29      30      31      32
0.8496002197790717413994 0.000000001141583333245 0.0595423933648906306138 0.0321984835864497082047
0.000000000000002220446 0.000000001141583333245 0.3457015569155127421297
33      34      35      36      37      38      39
0.0321984835864497082047 0.000000000000002220446 0.0595423933648906306138 0.0000000095267770211087
0.000000000000002220446 0.0595423933648906306138 0.8496002197790717413994
40      41      42      43      44      45      46
0.8496002197790717413994 0.000000000000002220446 0.0262210111496524837793 0.0321984835864497082047
0.5013604745398737572160 0.000000002766827167297 0.000000002766827167297
47      48      49      50      51      52      53
0.8496002197790717413994 0.0139524017064684060369 0.0000000029581759100971 0.0595423933648906306138
0.3457015569155127421297 0.1314046037796580967960 0.0262210111496524837793
54      55      56      57      58      59      60
0.000000002766827167297 0.4036640213243484209826 0.000000001141583333245 0.000000001176749447137
0.3457015569155127421297 0.0000000181294212544800 0.0595423933648906306138
61      62      63      64      65      66      67
0.000000000000002220446 0.000000001141583333245 0.000000000000002220446 0.0595423933648906306138
0.000000004855227579430 0.8496002197790717413994 0.0595423933648906306138
68      69      70      71      72      73      74
0.4036640213243484209826 0.000000002766827167297 0.000000000000002220446 0.0595423933648906306138
0.000000000000002220446 0.0595423933648906306138 0.0262210111496524837793
75      76      77      78      79      80      81
0.000000004855227579430 0.0000000469786253913346 0.000000001141583333245 0.000000005998873117498
0.000000002766827167297 0.000000002766827167297 0.000000000000002220446
82      83      84
0.000000002766827167297 0.000000001141583333245 0.000000002766827167297 0.0262210111496524837793
0.0000000002551359444929 0.4036640213243484209826 0.4036640213243484209826
89      90      91      92      93      94      95
0.0595423933648906306138 0.000000001176749447137 0.4036640213243484209826 0.0262210111496524837793
0.8496002197790717413994 0.0595423933648906306138 0.000000000000002220446
96      97      98      99      100      101      102
0.0595423933648906306138 0.4036640213243484209826 0.000000000000002220446 0.4036640213243484209826
0.5013604745398737572160 0.3457015569155127421297 0.000000004855227579430
103      104      105      106      107      108      109
0.4036640213243484209826 0.0595423933648906306138 0.4036640213243484209826 0.9148930601221268998913
0.5013604745398737572160 0.8496002197790717413994 0.8496002197790717413994
110      111      112      113      114      115      116
0.9148930601221268998913 0.8496002197790717413994 0.4036640213243484209826 0.8496002197790717413994
0.8496002197790717413994 0.8496002197790717413994 0.8496002197790717413994
117      118      119      120      121      122      123
0.9148930601221268998913 0.8496002197790717413994 0.9148930601221268998913 0.4036640213243484209826
0.8496002197790717413994 0.8496002197790717413994 0.9148930601221268998913
124      125      126      127      128      129      130
0.4036640213243484209826 0.5013604745398737572160 0.9148930601221268998913 0.5013604745398737572160
0.9148930601221268998913 0.9148930601221268998913 0.8496002197790717413994
131      132      133      134      135      136      137
0.5013604745398737572160 0.9148930601221268998913 0.3457015569155127421297 0.3457015569155127421297
0.000000005998873117498 0.000000002766827167297 0.000000002766827167297
138      139      140      141      142      143      144
0.000000000000002220446 0.000000002766827167297 0.000000001141583333245 0.000000002766827167297
0.0262210111496524837793 0.0262210111496524837793 0.0321984835864497082047
145      146      147      148      149      150
0.0595423933648906306138 0.000000001141583333245 0.0262210111496524837793 0.0000000095267770211087
0.5013604745398737572160 0.3457015569155127421297 0.3457015569155127421297
0.8496002197790717413994 0.4036640213243484209826 0.4036640213243484209826

```

## REGISTRO FOTOGRAFICO





