UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO

FACULTAD DE CIENCIAS HISTÓRICO SOCIALES Y EDUCACIÓN

UNIDAD DE POSGRADO

PROGRAMA DE MAESTRÍA EN CIENCIAS DE LA EDUCACIÓN

TESIS

Modelo de regresión logística en la predicción de la conducta antisocial en adolescentes de una Institución Educativa de la ciudad de Lambayeque. 2017.

Presentada para obtener el Grado Académico de Maestra en Ciencias de la Educación con mención en Psicopedagogía Cognitiva

Maestra: Loaiza Chumacero, Sandra Cecilia

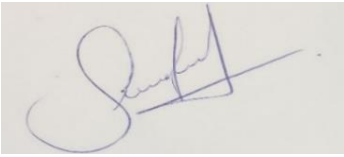
Asesor: Bobadilla Ocaña, Santiago

Lambayeque- Perú

2021

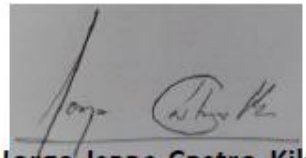
Modelo de regresión logística en la predicción de la conducta antisocial en adolescentes de una Institución Educativa de la ciudad de Lambayeque 2017.

Tesis presentada para obtener el Grado Académico de Maestra en Ciencias de la Educación con mención en Psicopedagogía Cognitiva.



Sandra Cecilia Loaiza Chumacero

Maestra



Dr. Jorge Isaac Castro Kikuchi

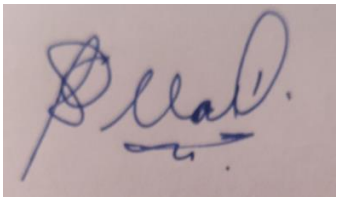
Presidente

Dra. María Del Pilar Fernández Celis

Secretario

Dra. Raquel Yovana Tello Flores

Vocal



Dr. SANTIAGO BOBADILLA OCAÑA

Asesor

DEDICATORIA

A mi Madre Mirian Chumacero Recoba

A mi hijo David Alonso mi hermoso niño por acompañarme en todo momento.

A mis Abuelos Ceferino Chumacero e Ida Recoba

AGRADECIMIENTOS

Mi más sincero agradecimiento al Dr. Eduar Vásquez Sánchez a la Dra. Gloria Ortiz Basauri por siempre confiar en mí, desde que era una estudiante universitaria.

A mi asesor Dr. Santiago Bobadilla Ocaña por apoyarme con sus conocimientos y por la confianza en esta investigación.

A mi querido profesor Jesús Salinas Flores por su asesoría en el análisis estadístico por compartir sus conocimientos.

A grandes profesionales de la Institución donde laboro:
Ing. María Aranguri.
Dr. Miguel Falla Rosado
Dra. Sofía Lavado Huarcaya.
Mgs. Jorge Fupuy

Por su apoyo en la investigación: Ing. Carlos Balarezo.

A los Miembros de la Policía Nacional de la Comisaría de Lambayeque:

Brigadier. Juan Ricardo Córdova Prado
Brigadier. Joel Calderón Vallejos
Brigadier. Mariella Capuñay Castillo

ÍNDICE

Índice de tablas **9**
Índice de figuras **10**
Resumen/abstract **11**
Introducción **12**

Capitulo I. Diseño teórico **17**
1.1. Definición de conducta antisocial 17
1.2. Factores de Riesgo y Protección 18
1.2. 1. Factores de riesgo asociados a adolescentes escolares 21
1.3. Educación Básica Regular 22
1.4 El Modelo de Regresión Logística 23
1.4.1. Introducción 23
1.4.2. El Modelo multivariante de Regresión Logística binaria 25

Capitulo II. Métodos y materiales **31**
2.1. Tipo de Investigación 31
2.2. Hipótesis 31
2.3. Operacionalización de Variables 31
2.4. Diseño de la Investigación 32
2.5 Población y Muestra 33
2.6. Métodos, Técnicas e Instrumentos de Recolección de Datos 33
2.7. Métodos y Procedimientos para La Recolección De Datos 34

Capitulo III. Resultados y discusión **35**
3.1. Resultados 35
3.1.1. Análisis exploratorio univariado 35
3.1.2. Aplicación de la Regresión Logística 46
3.2. Discusión 52

Capitulo IV. Conclusiones **54**

Capitulo V. Recomendaciones.....55

Bibliografía referenciada.....56

Anexos…….61

Anexo 1 Regresión logística en R……. ….61

Anexos 2. Fotografías de los Programas que realiza la Policía Nacional de la Oficina de Participación Ciudadana, con los estudiantes de la Institución Educativa…….66

Anexos 3. Informes de similitud del software Turnitin …….67

ÍNDICE DE TABLAS

Tabla 1:	Escala de evaluación del comportamiento…….	22
Tabla 2:	Matriz de Confusión ……..	30
Tabla 3:	Variable dependiente……..	31
Tabla 4:	Variables Independientes (Sociodemográficos)…….	32
Tabla 5:	Medidas Descriptivas de la Edad de los estudiantes de la Institución Educativa de Lambayeque, Julio 2018……	38
Tabla 6:	Medidas descriptivas del Rendimiento en Matemática y Comunicación Integral de los Estudiantes de una Institución Educativa de la ciudad de Lambayeque, Julio 2018……..	43
Tabla 7:	Prueba Binomial de la Variable Conducta Antisocial ……..	45
Tabla 8:	Variables finales para el modelamiento……..	46
Tabla 9:	Balanceando los datos con el Algoritmo SMOTE……..	48
Tabla 10:	Regresión logística con balanceo mediante SMOTE…….	49
Tabla 11:	Tabla Cruzada de Regresión Logística con SMOTE……..	49
Tabla 12:	Indicadores de clasificación de una Regresión Logística aplicando SMOTE……...	50
Tabla 13:	Predicción de nuevos estudiantes ……	51

ÍNDICE DE FIGURAS

Gráfico 1:	Función Logística …….	26
Gráfico 2:	Diseño de la Investigación……	32
Gráfico 3:	Sexo de los estudiantes de la Institución Educativa de Lambayeque, 2018…….....	35
Gráfico 4:	Grado de estudios de los estudiantes de la Institución Educativa de Lambayeque, Julio 2018…….	36
Gráfico 5	Edad de los estudiantes de la Institución Educativa de Lambayeque, Julio 2018……..	37
Gráfico 6	Gráfico de Caja de la Edad de los estudiantes de la Institución Educativa de Lambayeque, Julio 2018……..	38
Gráfico 7	Convivencia de los estudiantes de la Institución Educativa de Lambayeque, Julio 2018……..	39
Gráfico 8	Actualmente trabaja según los estudiantes de la Institución Educativa de Lambayeque, Julio 2018……..……	40
Gráfico 9	Rendimiento en Matemática de los estudiantes de la Institución Educativa de Lambayeque, Julio 2018……..	41
Gráfico 10	Rendimiento en Comunicación Integral de los estudiantes de la Institución Educativa de Lambayeque, Julio 2018……	42
Gráfico 11	Gráfico de caja del Rendimiento en Matemática y Comunicación Integral de los Estudiantes de una Institución Educativa de la ciudad de Lambayeque, Julio 2018…….	43
Gráfico 12	Selección de muestras: entrenamiento y prueba……...	47
Gráfico 13	Áreas bajo la curva (AUC)……	50

RESUMEN

La presente investigación es de enfoque cuantitativo de diseño no experimental, transversal correlacional. Se analizaron las variables sociodemográficas como: grado, sexo, edad, convivencia, trabaja, rendimiento en los cursos de matemática y comunicación, variables de comportamiento (conducta) en el aula, permitirán reconocer y pronosticar a los estudiantes con riesgo de conducta antisocial, mediante de un modelo de regresión logística binario. La muestra está conformada por 524 alumnos matriculados en el presente año 2018 del cual se obtuvo la información completa del registro de tutoría. Se concluyó que Aplicando Regresión Logística con SMOTE posee mejor desempeño como clasificador ante el riesgo de conducta antisocial; produce un alto índice de Sensibilidad, que es para predecir la categoría de Si está en Riesgo de Conducta Antisocial. Con respecto al resultado de la curva ROC, el área bajo la curva es de 84.25% clasifique correctamente a los estudiantes.

Palabras clave: Regresión logística; conducta antisocial.

ABSTRACT

The present research has a quantitative approach with a non-experimental, cross-sectional correlational design. Sociodemographic variables such as: grade, sex, age, coexistence, work, performance in mathematics and communication courses, behavioral variables (conduct) in the classroom, will allow to recognize and predict students at risk of antisocial behavior, by means of a binary logistic regression model. The sample consists of 524 students enrolled in the current year 2018 from which the complete information was obtained from the tutoring record. It was concluded that Applying Logistic Regression with SMOTE possesses better performance as a classifier before the risk of antisocial behavior; it produces a high index of Sensitivity, which is to predict the category of If at Risk of Antisocial Behavior. Regarding the result of the ROC curve, the area under the curve is 84.25% to correctly classify the students.

Keywords: Logistic regression; antisocial behavior.

INTRODUCCIÓN

La presente investigación constituye el desarrollo de un modelo para predecir que adolescente estudiante de educación Básica Regular del nivel secundario para identificar a los estudiantes con mayor probabilidad de estar en riesgo de una conducta antisocial.

Los datos a usarse en esta investigación provienen de Institución Educativa Emblemática "Juan Manuel Iturregui", en la ciudad de Lambayeque, el área de coordinación de tutoría nos proporcionó las fichas de los estudiantes las cuales se encontraba en formato físico digitalizando todas las fichas durante el año 2017, obteniendo una muestra de 524 estudiantes, de los cuales 21 están en riesgo de conducta antisocial y 503 no están en riesgo.

El estudio nace como una necesidad de focalizar la atención en aquellos estudiantes que necesitan el apoyo de la institución mediante el área de tutoría , la utilización de este modelo a partir de las variables sociodemográficas como: grado, sexo , edad, convivencia , trabaja, rendimiento en matemática y comunicación , variables de comportamiento (conducta) en el aula , permitirá que el modelo de regresión logística binaria identifique y prediga a los estudiantes que están en riesgo de comportamiento antisocial. .

Para la aplicación de la Regresión Logística se utilizaron los softwares: R Studio, Spss v. 25.0 para Windows y Excel. En términos generales el objetivo del estudio es conocer los factores que influyen en la conducta antisocial en adolescentes.

Por lo expuesto anteriormente, se planteó el siguiente problema de investigación ¿Al utilizar el modelo de regresión logística binaria para determinar el modelo explicativo, es posible predecir y clasificar qué adolescentes están en riesgo de comportamiento antisocial?

Teniéndose por objetivo general: Determinar un modelo predictivo aplicando la Regresión Logística para identificar los factores de riesgo y protección conducta antisocial en adolescentes de una Institución Educativa de la ciudad de Lambayeque; y como objetivos específicos: Identificar las variables predictoras de la conducta antisocial aplicando Regresión Logística Binaria.

Determinar un modelo que permita predecir y clasificar a estudiantes de conducta antisocial aplicando Regresión Logística Binaria.

El comportamiento antisocial es un problema que tiene graves consecuencias para los niños y los jóvenes. Los menores que exhiben un comportamiento antisocial generalmente exhiben un comportamiento agresivo repetido, robo, vandalismo y violaciones graves de las reglas en el hogar y la escuela (Rivera & Cahuana Cuentas, 2016).

Los principales factores de riesgo temprano para el comportamiento antisocial incluyen impulsividad, retraso mental y bajo rendimiento académico, mala supervisión de los padres, abuso físico infantil, disciplina punitiva o inestable de los padres, indiferencia de los padres, conflicto de los padres, ansiedad familiar, padres antisociales, escala grande familias, bajos ingresos familiares, compañeros antisociales, escuelas con altos índices de criminalidad y comunidades con altos índices de criminalidad (Farrington, 2005). Existen al menos tres vías importantes a través de las cuales los niños y adolescentes pueden desarrollar comportamientos antisociales graves (Frick & Viding, 2009).

Se hace una distinción entre factores globales (por ejemplo, estatus socioeconómico familiar), remotos (características de los padres), antecedentes (características familiares) y de proximidad (interacción entre padres e hijos) que operan dentro de la familia (Deković et al., 2003). El mecanismo causal que vincula estos factores de riesgo con resultados antisociales aún no está claro, la principal implicación de la intervención es que se debe implementar un plan que se concentre en estos factores de riesgo clave, especialmente un plan comunitario de múltiples componentes (Farrington, 2005).

Al delinear estas distintas vías hacia el comportamiento antisocial, hemos intentado ilustrar algunos conceptos clave de la psicopatología del desarrollo como la equifinalidad y la multifinalidad, la importancia de comprender la interfaz entre el desarrollo normal y anormal y la importancia de utilizar múltiples niveles de análisis para avanzar en la causalidad (Frick & Viding, 2009). Según el modelo hipotético de mediación, la influencia de factores remotos y de fondo parece ser mayoritariamente indirecta: después de considerar su asociación con

factores proximales, estos factores ya no están significativamente relacionados con la conducta antisocial. Discutió el impacto de estos hallazgos en el desarrollo de intervenciones adecuadas para los adolescentes y sus familias (Deković et al., 2003).

El comportamiento antisocial temprano proviene de problemas de comportamiento en la niñez, especialmente aquellos que se caracterizan por comportamientos agresivos y destructivos. Los defectos de autorregulación en múltiples áreas funcionales, desde la fisiología hasta la cognición, están relacionados con problemas de comportamiento tempranos y pueden hacer que los niños tengan un mayor riesgo de comportamiento antisocial en el futuro (Calkins & Keane, 2009).

Un grupo de jóvenes muestra un comportamiento antisocial que comienza en la adolescencia y dos grupos muestran un comportamiento antisocial que comienza en la niñez, pero que difieren en la presencia o ausencia de rasgos insensibles y sin emociones (Frick & Viding, 2009). Los resultados muestran que los factores proximales son predictores importantes del comportamiento antisocial, independientemente de su variación común con otros factores (Deković et al., 2003).

El comportamiento y la función de los padres también se han estudiado como predictores de trayectorias de comportamiento problemático temprano, y algunos datos apoyan la interacción de la crianza y la autorregulación como un predictor importante de patrones de comportamiento problemático y problemas continuos de regulación del comportamiento (Calkins & Keane, 2009). La relación con los compañeros también afecta y se ve afectada por las habilidades tempranas de autorregulación, las cuales pueden desempeñar un papel en el rendimiento y el éxito académicos posterior. Estos datos proporcionan evidencia de que el trasfondo social de las primeras relaciones familiares y con los compañeros es un regulador importante del mecanismo de autorregulación más cercano, y tanto los comportamientos sociales como los biológicos pueden involucrar tendencias antisociales tempranas (Calkins & Keane, 2009).

Las aplicaciones de este modelo predictivos de conducta antisocial nos van a permitir realizar intervenciones de apoyo temprano al adolescente (Justicia et al., 2006). Con esta información el área de tutoría de la Institución Educativa podrá desarrollar intervenciones focalizadas en aquellos adolescentes en riesgo y así poder realizar programas de prevención sobre niños y adolescentes (Darling-Hammond et al., 2020).

Es en este contexto que se ha propuesto el trabajo de investigación actual, cuyo principal objetivo se centrará en categorizar y predecir a los jóvenes en riesgo de conducta antisocial mediante la determinación de los patrones que determinan la situación de los estudiantes.

Para Walker et al., (1998) "Con una inversión de 35 dólares por alumno, el programa First Step to Success logró disminuir los niveles de agresión, y aumentar el nivel de atención en clase y el rendimiento académico en niños cuyas habilidades sociales, de control emocional y de razonamiento moral fueron desarrolladas, frente a un grupo de niños que no fue parte del programa" (Walker et al., 1998).

Los resultados mostraron que la participación en actividades de ocio altamente estructuradas se asoció con bajos niveles de comportamiento antisocial, mientras que la participación en actividades de baja estructura (es decir, un centro de recreación juvenil) se asoció con altos niveles de comportamiento antisocial. La combinación de participación en una actividad no estructurada y la falta de una participación altamente estructurada pareció ser particularmente problemática para el comportamiento antisocial en los niños (Mahoney & Stattin, 2000).

Un mensaje importante es que las intervenciones y el apoyo conductual positivo son inclusivos en toda la escuela, y la implementación exitosa enfatiza la identificación, adopción y uso sostenible de políticas, sistemas, toma de decisiones y prácticas efectivas basadas en datos. También se analizan los desafíos a nivel del sistema (Sugai et al., 2000).

Los participantes en actividades no estructuradas también se caracterizaron por relaciones desviadas con los compañeros, malas relaciones entre padres e hijos y recibieron poco apoyo

de su líder de actividad en comparación con los adolescentes que participaban en actividades comunitarias más estructuradas (Mahoney & Stattin, 2000).

I. CAPITULO I: DISEÑO TEÓRICO

1.1. DEFINICIÓN DE CONDUCTA ANTISOCIAL

El comportamiento antisocial se refiere a una serie de comportamientos heterogéneos fuera de las normas, reglas o leyes del grupo social desarrollado por el sujeto, como ataques personales, robos y violaciones de las reglas sociales (Acquaviva et al., 2018).

Según Coelho, (2020), El comportamiento antisocial se refiere a una combinación de comportamientos, acciones o prácticas inapropiadas que promueven el caos social y violan o invaden la ley; los discursos que realizan estos mismos comportamientos también se denominan comportamientos antisociales. El autor cree que el comportamiento antisocial se puede dividir en tres factores, la severidad de la evaluación, la distancia entre la evaluación y las normas sociales establecidas, y el contexto sociocultural en el que se implementa, lo que ayuda a evaluar y distinguir comportamiento antisocial (Coelho et al., 2020).

Calkins, (2009), nos menciona que la conducta antisocial hace referencia a situaciones que incumplen las normas sociales y derechos de los demás. La definición de conducta antisocial es difícil de limitarlo ya que hay distintas maneras de apreciarlas, además la conducta antisocial hace referencia al contexto sociocultural en que se desarrolla una conducta (Calkins & Keane, 2009).

Gershoff (2002) el cual nos hace referencia que El comportamiento antisocial se basa en la desobediencia o violación de las normas sociales o culturales. Los que producen un comportamiento antisocial se destacan o se caracterizan por la falta de adaptación, el egocentrismo, la impulsividad, la irresponsabilidad, el autoritarismo, la conciencia moral y ética subdesarrollada, la disciplina, falta de tolerancia a los contratiempos y falta de respeto a la integridad de los compañeros. Las personas que exhiben este comportamiento carecen del superyó en psicoanálisis, que está relacionado con el crimen, la psicosis o el comportamiento antisocial (Gershoff, 2002).

Olmedo Hernández et al., (2019, nos dice que la conducta antisocial se divide en dos tipos primordiales, estas son, la agresión, el cual busca hacer daño o perjudicar de manera física o verbal a su prójimo causándole dolor de forma directa e indirectamente; y la conducta furtiva, este tipo de conducta no tiene como propósito dañar o afectar a los sujetos de manera directa, sino que afectara o tendrá algún efecto según las acciones que el realice (Olmedo Hernández et al., 2019).

El comportamiento antisocial también se desarrolla a través de la interacción social dentro de la familia y la comunidad. Afecta continuamente el temperamento, la capacidad cognitiva y la interacción negativa con los compañeros del niño, y afecta en gran medida la capacidad del niño para resolver problemas de manera cooperativa. Muchas personas también marcan comportamientos que se considera que violan los códigos de conducta sociales actuales como comportamientos antisociales (Rivera & Cahuana Cuentas, 2016).

1.2. FACTORES DE RIESGO Y PROTECCION

- **FACTORES DE RIESGO**

Pedrosa et al, (2020) plantea que los factores de riesgo es cualquier acto o circunstancia en la que se encuentre expuesto un individuo o grupo de individuos. Al sumarse estos factores de riesgo (ambientales, familiares, psicológicos, etc.) pueden lograr aislar a un individuo teniendo como efecto un fenómeno de interacción (Pedrosa et al., 2020). Según de los Ángeles, (2011) nos informa que dentro del universo adolescente existen gran cantidad de factores, las cuales estarían llevando que los adolescentes adquieran estas conductas de riesgo e influenciando. Estos factores tienen como origen el ámbito interno, externo y estarían relacionados con aspectos personales, familiares y sociales (de los Ángeles Páramo, 2011).

Los delitos violentos son una de las principales causas de inestabilidad social, lesiones y muerte en los países de ingresos bajos y medianos (Murray et al., 2018). Comprender las causas del comportamiento antisocial (es decir, violencia, crimen, violación de las reglas) es esencial para desarrollar estrategias de prevención e intervención más efectivas

(Gard et al., 2019). Se han hecho mayores esfuerzos para combinar factores biológicos y sociales para explicar el desarrollo antisocial (van Hazebroek et al., 2019). Los estudios longitudinales en países de ingresos altos han proporcionado evidencia importante sobre los precursores del desarrollo de la violencia y otras conductas antisociales (Murray et al., 2018).

Para reducir el comportamiento antisocial y los problemas individuales y sociales asociados, se justifica la comprensión de los determinantes de comportamiento antisocial (van Hazebroek et al., 2019). Sin embargo, puede haber influencias únicas o diferentes efectos de factores de riesgo en otros entornos sociales (Murray et al., 2018). En el contexto del riesgo social, la falta de excitación se asoció con la agresión proactiva, mientras que la sobreexcitación se asoció con la agresión reactiva (van Hazebroek et al., 2019).

Muchos factores de riesgo tienen aproximadamente los mismos efectos promedio que cuando se estudian en países de ingresos altos. La estabilidad de la agresión durante un período de tres años es casi idéntica en los países de ingresos bajos y medios y en los países de ingresos altos (Murray et al., 2018). Las asociaciones entre la interacción biosocial y comportamiento antisocial fueron más fuertes para los tipos de comportamiento antisocial más severos y violentos (van Hazebroek et al., 2019). Las dimensiones de la psicopatología comórbida, como el bajo autocontrol, la hiperactividad y la búsqueda de sensaciones, están asociadas con el comportamiento antisocial en los países de ingresos bajos y medios, pero algunos factores de salud física tempranos tienen efectos consistentemente débiles o nulos (Murray et al., 2018). Los hallazgos empíricos se discuten en términos de distintas perspectivas teóricas biosociales que tienen como objetivo explicar comportamiento antisocial y se describen importantes cuestiones empíricas no resueltas (van Hazebroek et al., 2019).

- **FACTORES DE PROTECCION**

Los factores de protección no son necesariamente actos beneficiosos o circunstancia que buscan el bien para el sujeto ante una situación de riesgo. Sino que en estos últimos años el concepto de factores de protección hace referencia a resultados positivos,

independientemente que exista o no una situación de alto riesgo, argumentado por algunos investigadores es aceptable decir protección siempre que se esté interactuando con una situación de riesgo y podamos predecirla (Office of the Surgeon General, 2001). Los recursos mas importantes con los que cuenta un adolescente resiliente son: mantienen una buena relacion con sus padres, un ambiente escolar calido y abierto, apoyo social, una buena autoestima, buen manejo de afrontamiento ante situaciones complicadas, entre otros, a si mismo se ha detectado que la comunicación y la relaciones familiares estan asociados afactores de proteccion y llevar una vida saludable (Uriarte Arciniega, 2005).

Los resultados de la investigación también muestran que los comportamientos altamente pro-sociales y la regulación emocional de los compañeros (pero no de los padres) debilitan la conexión entre la violencia del vecindario y el comportamiento antisocial (Criss et al., 2017). Los resultados empíricos muestran varios candidatos para promover o amortiguar la protección: variables individuales (capacidad cognitiva, procesamiento de información, temperamento y características biológicas), variables familiares (vínculos emocionales, comportamiento parental apropiado, interés de los padres en la educación y situación socioeconómica); factores escolares (buenas calificaciones, relaciones escolares y un ambiente educativo positivo), factores de pares (relaciones con amigos que no se desvían, cierto aislamiento social) y factores comunitarios (una comunidad cohesionada de la que no se priva) (Lösel & Bender, 2017).

1.2.1. FACTORES DE RIESGO ASOCIADOS A ADOLECENTES ESCOLARES

- **FACTORES DE RIESGO AMBIENTALES**

  En este factor contextual o ambiental, donde se ve involucrado los medios de comunicación, grupos de iguales, instituciones, comunidades, agrupaciones, escuelas y otros ámbitos, que también se pueden observar, fomentar o alentar a

situaciones agresivas, violentas, o coléricas de algunos sujetos (Huesmann & Taylor, 2006).

La relación que existe entre las comunidades, barrios y la conducta antisocial Según De la Peña, (2005) nos hace referencia que una baja alineacion con el barrio y el desorden de la comunidad, el alcance a las drogas y armas de fuego, la manifestacion de violencia en el barrio y otros medios, el racismo e irregularidades de normas en las comunidades son factores de riesgos que influyen mucho en el adolescente, logrando asumir conductas antisociales en los individuos (Peña Fernández & Peña Fernández, 2005).

- **FACTORES DE RIESGO FAMILIARES**

Farrington, 2005 nos manifiesta en este factor que la crianza y la enseñanza familiar hacia los hijos son muy esenciales, el cual nos ayuda a apreciar el desarrollo y aparición de estas conductas antisociales. Experiencias como el divorcio, situaciones de violencia entre familiares, violaciones domestica pueden ser el principal motivo para el desarrollo de estas conductas (Farrington, 2005).

De la Peña, (2005) el nucleó familiar es el principal ambiente social donde las personas aprenden o adquieren valores y visiones e instaura sus primeras conductas. Con respecto a los tipos de comportamiento estudiados de los sujetos son consecuencia de las experiencias familiares, escolares, y otros llamados “problemáticos”, como son la agresión, psicopatologías y delincuencia, llamados conductas antisociales que ayudan a la fomentación del desorden (Peña Fernández & Peña Fernández, 2005).

- **FACTORES DE RIESGO ESCOLARES**

Nos platea que los factores de riesgo que repercuten en un adolescente escolar hacia una conducta antisocial o al abandono escolar son: integración al mercado

laboral, edad de los estudiantes, ocupación de los padres, los estudios de los padres, organización de la familia, numero de hermanos (van Hazebroek et al., 2019).

La ausencia de interés hacia la educación, el bajo rendimiento académico, y la deficiente calidad de las escuelas son parámetros o indicadores relacionados con la educación, se ven involucrados en el comportamiento antisocial y violento de los adolescentes escolares según lo mencionado por De la Peña, (2005).

1.3.EDUCACIÓN BASICA REGULAR (EBR) - EDUCACION SECUDARIA

- **CALIFICACION**

La escala de calificacion según el Ministerio de Educación, (2005), en educacion secudanria EBR, como minima nota aprobatoria es de 11; la calificacion del comportamiento es categorica, literal y descriptiva la cual es asignada por el tutor encargado, el auxiliar de apoyo y una persona asiganada para el fin. Teniendo como referencia los informes de los docentes de cada area, donde se refleja los casos mas relevantes del desempeño de los estudiantes y seran evaluados bimestral o trimestralmente.

Aspectos sugeridos para la evaluación del comportamiento	Comportamiento	Escala
- Puntualidad y asistencia - Presentación personal - Cuidado del Patrimonio institucional - Respeto a la propiedad ajena - Orden - Limpieza - Respeto a las normas de convivencia	Muy Bueno	AD
	Bueno	A
	Regular	B
	Deficiente	C

Tabla 1. Escala de evaluación del comportamiento

Fuente: Ministerio de Educación 2005

- **MERITO Y ESTIMULOS**

  Según el Ministerio de Educación, (2005), se estimula y se hace reconocimiento a los estudiantes que han finalizado cada grado y al concluir su educación secundaria EBR en:

  - Rendimiento académico
  - Comportamiento
  - Éxitos sobresalientes en las asignaturas como: ciencias, humanidades, educación física, arte, deportes y otros.

1.4. MODELO DE REGRESIÓN LOGÍSTICA

1.4.1 Introducción

El objetivo de la investigación se centra en predecir el comportamiento de una variable categórica con dos grupos: ser un adolescente con riesgo de conducta antisocial o no (Vega-Vilca & Guzman, 2011). El problema de discriminación se convierte en prever el valor de la variable ficticia Y en un nuevo elemento del que conocemos variables X. Si el valor previsto está más próximo a cero que a uno, se clasificará para el elemento en la primera población (Díaz Rosas, 1995). Para modelar este tipo de relaciones se utilizan los modelos de respuesta cualitativa. El modelo de esta clase más utilizado es el modelo logístico. Este modelo requiere realizar menos supuestos, lo que permite obtener resultados más robustos y es flexible en cuanto a la naturaleza de las variables explicativas, pues éstas pueden ser de escala y categóricas. Permite estudiar el impacto que tiene cada una de las variables independientes en la probabilidad de que ocurra el suceso de estudio (Jaume & Morar-Catalá, 2001).

La técnica de la regresión logística se origina en la década de los años 60 con el trabajo de Cornfield, Gordon y Smith (1961). Walter y Duncan (1967) ya la utilizan en la forma actual, siendo a partir de los años 80, con la ayuda de la informática aplicada, que se generaliza su uso (Domínguez Alonso & Aldana Padilla, 2001). La regresión logística mezcla dos tradiciones del análisis

estadístico: el análisis de tablas de contingencia con el tratamiento de modelos log-lineales, y el análisis de regresión por mínimos cuadrados ordinarios. En ambos casos nos encontramos con limitaciones que la regresión logística resuelve: en el primer caso los modelos de dependencia no podían utilizar variables continuas y en el segundo las variables categóricas no siempre funcionan como buenos predictores (López-Roldán & Fachelli, 2018).

El modelo de regresión logística permite estimar la probabilidad de un suceso que depende de los valores de ciertas covariables. Supongamos que un suceso (o evento) de interés A; Si está en riego puede presentarse o No en cada uno de los individuos de una cierta población. Consideremos una variable binaria y que toma los valores:

y = 1 si A, y = 0 no A

La probabilidad de A no depende de otras variables, significa que P (A) = p; la racionalidad de una sola observación es

pues :
L = p si y = 1; L = 1 - p si y = 0:

Si realizamos n pruebas independientes y observamos y1,…… yn, la verosimilitud es

$$L = \prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i} = p^k(1-p)^{n-k}$$

siendo $k = \sum yi$ la frecuencia absoluta de A en las n pruebas. Para estimar p resolvemos la ecuación de verosimilitud

$$\frac{\partial}{\partial p} \ln L = 0$$

cuya solución es $\hat{p} = k/n$, la frecuencia relativa del suceso A. La distribución asintótica de $\hat{p}$ es normal N (p, p (1- p)/n).

El proceso de análisis de la regresión logística se puede dividir en varias etapas o tareas:

1) **Selección de las variables del modelo:** identificando las variables que aportan al modelo las cuáles son razonables en el contexto de los objetivos de la investigación.(Calduch Cervera, 2012).

2) Estimación del coeficiente de la variable independiente: Se puede utilizar un marco de probabilidad llamado estimación de máxima verosimilitud para estimar los parámetros de un modelo de regresión logística.

3) Clasificación de casos: La clasificación de casos se basa en la probabilidad predicha de que el caso sea un evento (el valor máximo de la variable dependiente (DV), calculado utilizando la ecuación del modelo actual).

4) Análisis de los residuos: Utilizando el modelo seleccionado, se puede realizar un análisis más detallado de los residuales para detectar la existencia de casos extremos, es decir, cuando existe una diferencia significativa (una desviación de más de dos unidades) entre la probabilidad observada y la probabilidad predicha por el modelo. Y su eliminación puede mejorar el ajuste del modelo. (López-Roldán, P.; Fachelli, S. , 2016).

1.4.2. El Modelo multivariante de Regresión Logística binaria

El modelo multivariante de Regresión logística binaria, considera como variable respuesta a dos valores (dicotómica). En el caso de esta investigación, la variable de respuesta es si los estudiantes están en riesgo de comportamiento antisocial, se le asigna un valor de 1 y a los estudiantes que no están en riesgo se les asigna un valor de 0, están categorías son mutuamente excluyentes.

$$Y_i = \begin{cases} 1, \mathrm{P}(Y_i = 1) = P_i \\ 0, \mathrm{P}(Y_i = 0) = 1 - P_i \end{cases}$$

Su valor esperado

$$E(Y_i) = 1xP_i + 0x(1 - P_i) = P_i$$

Pensaremos que i Y se explica por la variable independiente i X_2 , i X 3 , …, ki X , designamos $_i Z$ como:

$$Z_i = \beta_1 + \beta_2 X_{2i} + \ldots\ldots + \beta_k X_{ki} = f \qquad X_{2i.}\ldots\ldots X_{ki} - \begin{bmatrix}\beta_1\end{bmatrix} \begin{bmatrix}\beta_2\end{bmatrix} \begin{bmatrix}\ldots\ldots\end{bmatrix} \begin{bmatrix}\ldots\ldots\end{bmatrix} \begin{bmatrix}\beta_k\end{bmatrix}$$

Entonces la esperanza de i Y dado las variables independientes, es:

$$E(Y_i / X_{2i}, \cdots, X_{ki}) = F(\beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}) = F(Z_i)$$

El modelo quedaría como:

$$Y_i = E(Y_i / X_{2i}, \cdots, X_{ki}) + U_i = F(Z_i) + U_i$$

Ui es una perturbación aleatoria.

Para nuestro ejemplo, usamos una función lógica similar a la función F para obtener el modelo logit:

$$P_i = E(Y_i / X_{2i}, \cdots, X_{ki}) + U_i = F(Z_i) = \frac{1}{1 + e^{-Z_i}} = \Lambda(Z_i)$$

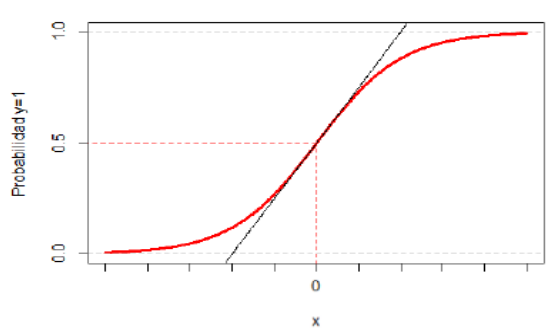


Gráfico 1. Función logística

a) Coeficientes de Regresión

Para comparar la significancia global en el modelo logit, usando estadísticas de razón de verosimilitud (RV). Por lo general, se utiliza una prueba completa (Reyes Rocabado et al., 2007).

En este caso, la hipótesis nula a contrastar es que todos los coeficientes de las variables independientes son cero (significancia del modelo), es decir, para determinar si al menos una variable independiente es significativa. Los siguientes supuestos:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \ldots\ldots\ldots\ldots \beta_k = 0$$

H_1: Al menos uno de los coeficientes es distinto de cero

En un modelo con un término independiente de 0, es decir, sin rechazar H0, se trata de tantas observaciones como admitir que un modelo que contiene solo constantes predice Y mejor que un modelo ajustado con n predictores. Por el contrario, si se rechaza o H0, esto indicará que al menos un coeficiente no es cero.

En la regresión logística, esto se realiza mediante la prueba G o la prueba de razón de verosimilitud. Estadístico de razón de verosimilitud G. Se trata de comparar todos los modelos producidos eliminando todas las covariables con el modelo completo.

En este caso cada estadístico G sigue una χ^2 (no se asume normalidad). La insignificancia significa que el modelo sin un predictor no se deteriorará en relación con el modelo completo (es decir, su existencia es irrelevante), por lo que de acuerdo con la estrategia de obtener el modelo más pequeño (el principio de parsimonia), significa que el predictor debe ser eliminado del modelo, porque no tiene ninguna contribución a él. La prueba no asume

ninguna distribución en particular, por lo que se recomienda estudiar la significancia del coeficiente.

G = D (Modelo sin las variables) – D (Modelo con las variables)

$$G = \frac{-2\ln \quad verosimilitud_del_modelo_solo_con_la_constante(Lo)}{verosimilitud_del_modelo_solo_seleccionado(Lp)}$$

Tiene una distribución chi-cuadrado con p -1 grados de libertad, donde p representa el número de parámetros en el modelo estudiado. Esta estadística se basa en la función de verosimilitud de cada modelo y finalmente compara la probabilidad de que los datos estimados por cada modelo representen el verdadero valor observado de la variable de respuesta (Hosmer et al., 2000).

b) Prueba de bondad de ajuste de Hosmer y Lemeshow

La prueba de Bondad de ajuste de Hosmer-Lemeshow Este indicador permite evaluar el modelo de forma global, es decir, qué tan bien las predicciones coinciden con el resultado real observado.

La prueba de Hosmer y Lemeshow se usa ampliamente en regresión logística. Ésta es una prueba de bondad de ajuste del modelo propuesto. Esta es una prueba en la que se evalúa la distancia entre la Oi observada y la Ei esperada.

$$HL = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \quad \text{, donde } E_i = n_i \bar{p}_i$$

Donde i n es el número de patrones de predicción del i-ésimo grupo, que es el número de respuestas positivas registradas para la variable de respuesta del i-ésimo patrón de predicción (Y = 1).

Donde:

Oi : número de unos en el decil i-ésimo

$\overline{p_i}$: media de probabilidades en el decil i-ésimo

ni : número de observaciones en el decil i-ésimo

Las hipótesis planteadas según la prueba son:

H_0: El modelo de regresión logística se ajusta a los datos.

H_1: El modelo de regresión logística no se ajusta a los datos.

"**Hosmer y Lemeshow" (HL)** La distribución es chi-cuadrado, el criterio de decisión será $\chi^2 \geq \chi^2_\alpha$, No rechace la hipótesis original y saque la conclusión de que el modelo se ajusta a los datos.

Si rechazamos "H_0", significa que el modelo ajustado no es suficiente.

c) Curva ROC

La curva característica de rendimiento del receptor o curva ROC es un gráfico que ilustra la capacidad de diagnóstico del sistema clasificador binario cuando cambia su umbral de discriminación. Cabe recordar que la diagonal principal es la línea correspondiente a la peor prueba diagnóstica, y el área debajo de ella es 0.5.

El diagnóstico puede explicarse por el área bajo la curva ROC, que se utiliza para medida de discriminación, y representa a todos los posibles pares de individuos compuestos por individuos con incidentes e individuos sin incidentes, el modelo predice la proporción de personas con mayor probabilidad de incidentes(Cerda & Cifuentes, 2012).

H_0: El área bajo la curva ROC es igual a cero.

H_1: El área bajo la curva ROC no es igual a 0.5

Si rechazamos la "H_0" asociado a un "p-value", implica que el modelo ajustado es el adecuado. A partir de un área de 0.7 la discriminación del modelo se considera aceptable.

El área bajo la curva ROC es siempre mayor o igual que 0.5: un valor que comienza desde 0,8 se considera una buena discriminación. Un valor de 0,9 se considerará muy bueno. Si el valor del área es 1, la discriminación será perfecta. Hosmer y Lemeshow (2000).

d) Matriz de Confusión

Una matriz de confusión es una tabla que se usa a menudo para describir el desempeño de un modelo de clasificación (o "clasificador") en un conjunto de datos de prueba con valores de verdad conocidos. La forma de la matriz de confusión es:

Tabla 2. Matriz de Confusión

Observado	**Pronóstico**	
	Clase 1	**Clase 2**
Clase 1	**A**	**B**
Clase 2	**C**	**D**

A es el número de observaciones de la clase 1 predichas correctamente

B es el número de observaciones de la clase 1 predichas incorrectamente

C es el número de observaciones de la clase 2 predichas incorrectamente

D es el número de observaciones de la clase 2 predichas correctamente.

$$Error\ Total = \frac{B + C}{n^{\circ}\ total\ observaciones}$$

Error de tipo I: Se calcula como:

$$Error\ tipo\ I = \frac{B}{A + B}$$

Error de tipo II: Se calcula como:

$$Error\ tipo\ II = \frac{C}{C + D}$$

II. CAPÍTULO II. MÉTODOS Y MATERIALES.

2.1. Tipo de Investigación

La presente investigación es de tipo correlacional, ya que se comprueba si existe alguna relación entre las variables independientes y la variable dependiente de manera individual, después se halla que variables influyen en la variable independiente (riesgo de conducta antisocial). Por otro lado, el estudio es del tipo predictivo ya que se obtuvo una regla que permita identificar si un adolescente con ciertas características tiene riesgo de presentar conducta antisocial.

2.2.Hipótesis

El modelo predictivo mediante factores la Regresión Logística me permitirá identificar los factores de riesgo y protección conducta antisocial en adolescentes de una Institución Educativa de la ciudad de Lambayeque.

2.3. Operacionalización de Variables

1. Variable dependiente

Se define como una variable de comportamiento antisocial, que se utiliza para identificar si un estudiante está en riesgo.

Tabla 3. Variable Independiente.

Tipo	Variable	Descripción	Valores	Etiqueta
Dicotómica	Situación	Riesgo de conducta antisocial	0	No está en riesgo
			1	Si está en riesgo

No riesgo (calificación en Conducta AD – Muy bueno, A – Bueno, B – Regular).

Estudiantes en Riesgo (Calificación en conducta C – Deficiente).

2. Variables Independientes

Tabla 4. Variables Independientes (Sociodemográficos)

Tipo	Variable	Descripción	Valores	Etiqueta
Categórica	Grado	año de estudios en educación secundaria (1° año, 4° año, 5° año	1 2 3	1° año 4° año 5° año
Dicotómica	Sexo	Sexo del estudiante	0 1	Femenino Masculino
Continua	Edad	Edad del estudiante	Número en años	
Categórica	Convivencia	Con quién vive actualmente	1 2 3 4	Mis padres Solo madre Solo padre Otros familiares
Dicotómica	Trabaja	Trabaja actualmente percibe un ingreso	0 1	No Si
Continua	Rendimiento en Matemática	Notas del curso de Matemática	Notas	0 - 20
Continua	Rendimiento en Comunicación	Notas del curso de Comunicación	Notas	0 - 20

2.4.Diseño de la Investigación

La investigación de diseño no experimental, puesto que su fin "es estudiar variables sin manipularlas, sino observar fenómenos en su estado natural para analizarla recolectando la información de la oficina de tutoría y de la secretaria académica de la Institución Educativa" (Hernández et al. 2014, p. 152).

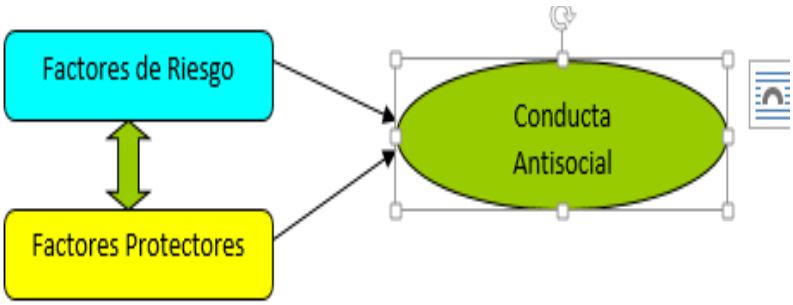


Gráfico 2. Diseño de la Investigación.

2.5. Población y Muestra

Considerando la población de estudio compuesta por 1640 alumnos, más las variables descritas en el punto anterior, se utilizó la metodología denominada análisis de eventos históricos, la cual estudia a los sujetos tomando en cuenta la unidad de tiempo en que son observados.

Para Allison (1984) un "evento" es un cambio que se produce en la vida de una persona o colectividad, producto de un quiebre entre lo que había antes y lo que vendrá después. Esta metodología, entonces, busca identificar qué variables son las que influyen en que dicho quiebre se produzca o no, controlando por una unidad de tiempo determinada. En este caso, la unidad de tiempo escogida es el año, debido a que corresponde al periodo lectivo mínimo existente para medir la conducta del adolescente los cuál esta información la brinda la coordinación de tutoría de la Institución Educativa.
El análisis de eventos históricos considera el caso observado en una unidad de tiempo determinada como una unidad de observación.

Muestra

La muestra está conformada por 524 alumnos matriculados en el presente año 2018 del cual se obtuvo la información completa del registro de tutoría.

2.6.Métodos, Técnicas E Instrumentos De Recolección De Datos

La técnica a utilizar es el de eventos históricos considera el caso observado en una unidad de tiempo, las variables las cuales están en el registro de los adolescentes con respecto al área académica en la oficina de Informática de la Institución educativa y la evaluación del comportamiento en el área de coordinación de tutoría.

Las fichas de recolección de datos del área de tutoría donde se registran las incidencias de los estudiantes como Presentación Personal, Puntualidad, Convivencia, Conservación de infraestructura.

2.7.Métodos y Procedimientos para la Recolección De Datos.

En primer lugar, se elaboró un reporte mediante la base de datos http://escale.minedu.gob.pe/, se pudo identificar que el centro educativo con mayor número de estudiantes matriculados en la ciudad de Lambayeque es la Institución Educativa Juan Manuel Iturruegui. Lo cual también se contactó con el subdirector y coordinador del Área de tutoría el cuál se le solitaria la información mediante oficio en las cuales se les explicara los objetivos, importancia y procedimientos de la investigación.

III. CAPITULO III. RESULTADOS Y DISCUSIÓN.

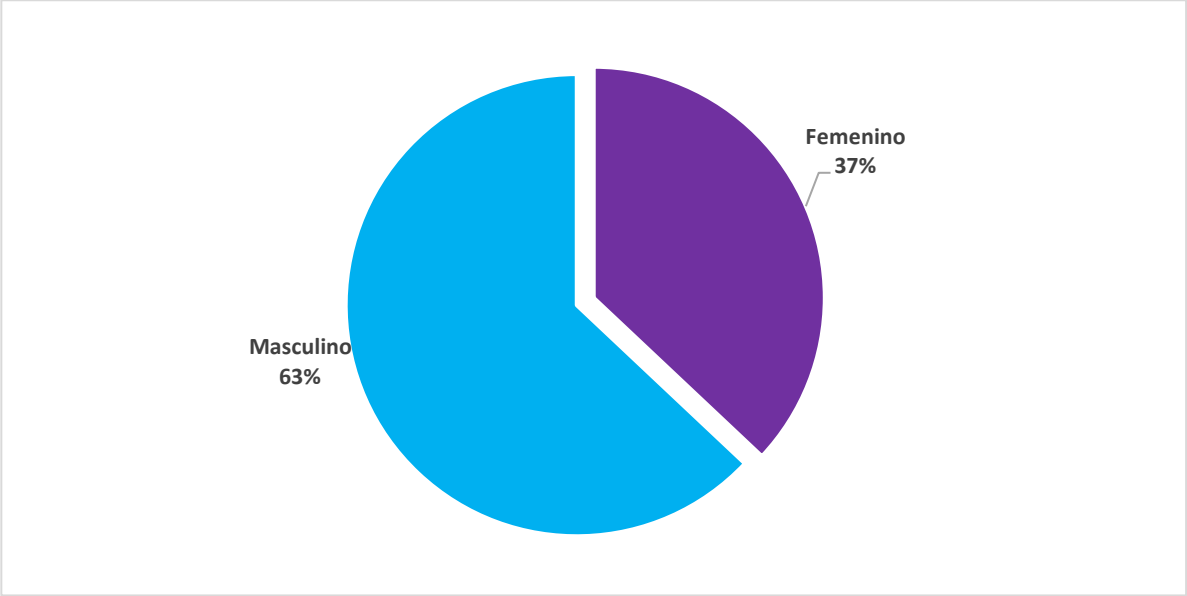
3.1.RESULTADOS

3.1.1 Análisis exploratorio univariado:

Se describieron las características sociodemográficas de los estudiantes matriculados en el año 2018.

A) Características Sociodemográficas de los estudiantes

Gráfico 3. Sexo de los estudiantes de la Institución Educativa de Lambayeque, 2018.

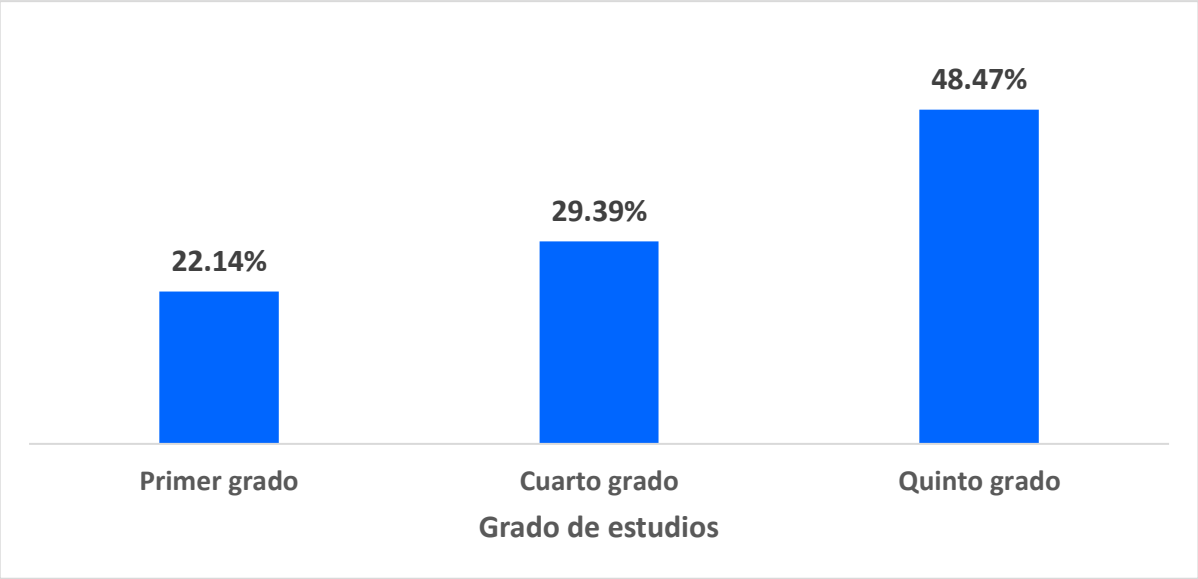


Fuente: Datos recolectados de la Oficina de Tutoría de la I. E. – Julio 2018.

Elaboración propia

Al realizar el análisis de las variables en estudio, el sexo de los estudiantes el 37% (femenino) y el 63% (masculino).

Gráfico 4. Grado de los estudiantes de la Institución Educativa de Lambayeque, Julio 2018.

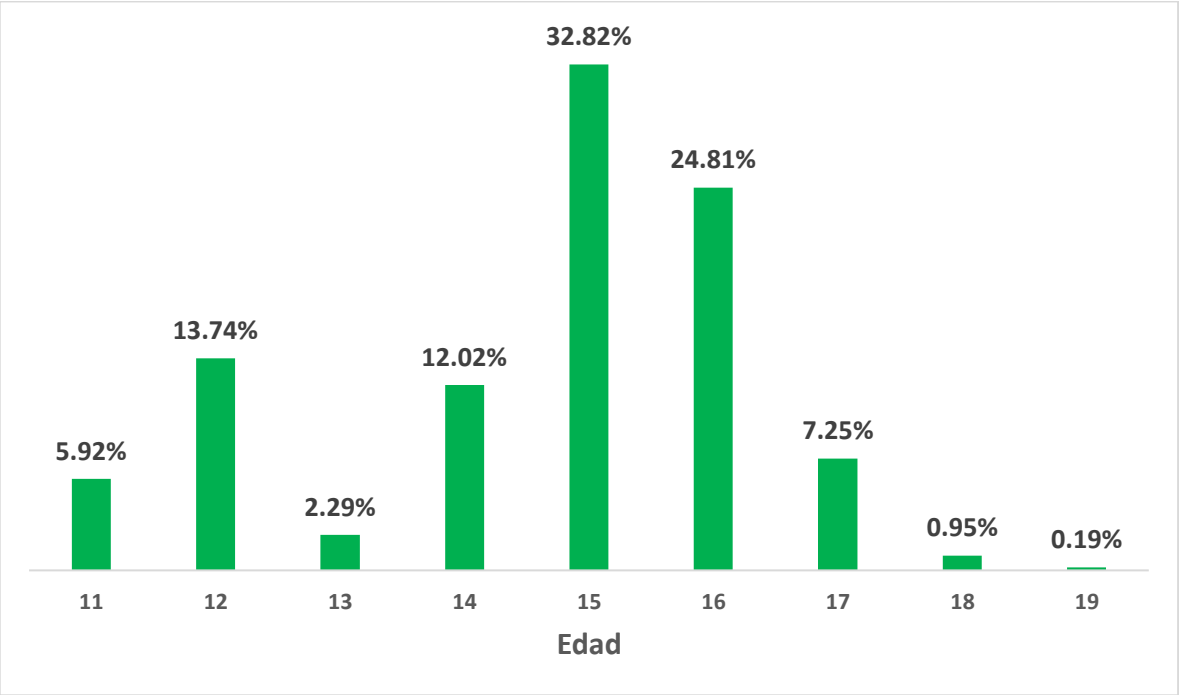


Fuente: Datos recolectados de la Oficina de Tutoría de la I. E. – Julio 2018.

Elaboración propia

Según el grado de estudios los estudiantes son del nivel secundario el 22.1% son del primer grado, 29.4% del segundo grado y el 48.5% del quinto grado de educación secundaria de la Institución Educativa.

Gráfico 5. Edad de los estudiantes de la Institución Educativa de Lambayeque, Julio 2018.



Fuente: Datos recolectados de la Oficina de Tutoría de la I. E. – Julio 2018.
Elaboración propia

Según las edades de los estudiantes de educación secundaria se observa que la edad mínima es de 11 años y la máxima es de 19 años bajo estudio. Así mismo se observa según el número de estudiantes la mayoría tiene 15 años (32.82%), 16 años (24.81%), y en menor número estudiante de18 (0.95%) y de 19 años (0.19%).

Tabla 5. Medidas Descriptivas de la Edad de los estudiantes de la Institución Educativa de Lambayeque, Julio 2018.

Medidas Descriptivas		Edad
Media		14.61
95% de intervalo de confianza para la media	Límite inferior	14.47
	Límite superior	14.76
Media recortada al 5%		14.67
Mediana		15.00
Varianza		2.907
Desviación estándar		1.705
Mínimo		11
Máximo		19
Rango		8
Rango intercuartil		2
Asimetría		-0.645
Curtosis		-0.388

Fuente: Datos recolectados de la Oficina de Tutoría de la I. E. – Julio 2018.
Elaboración propia

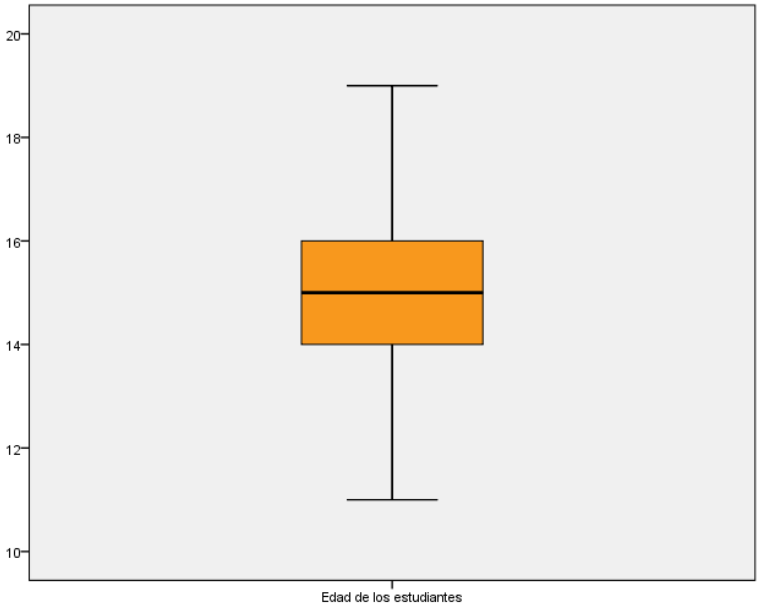
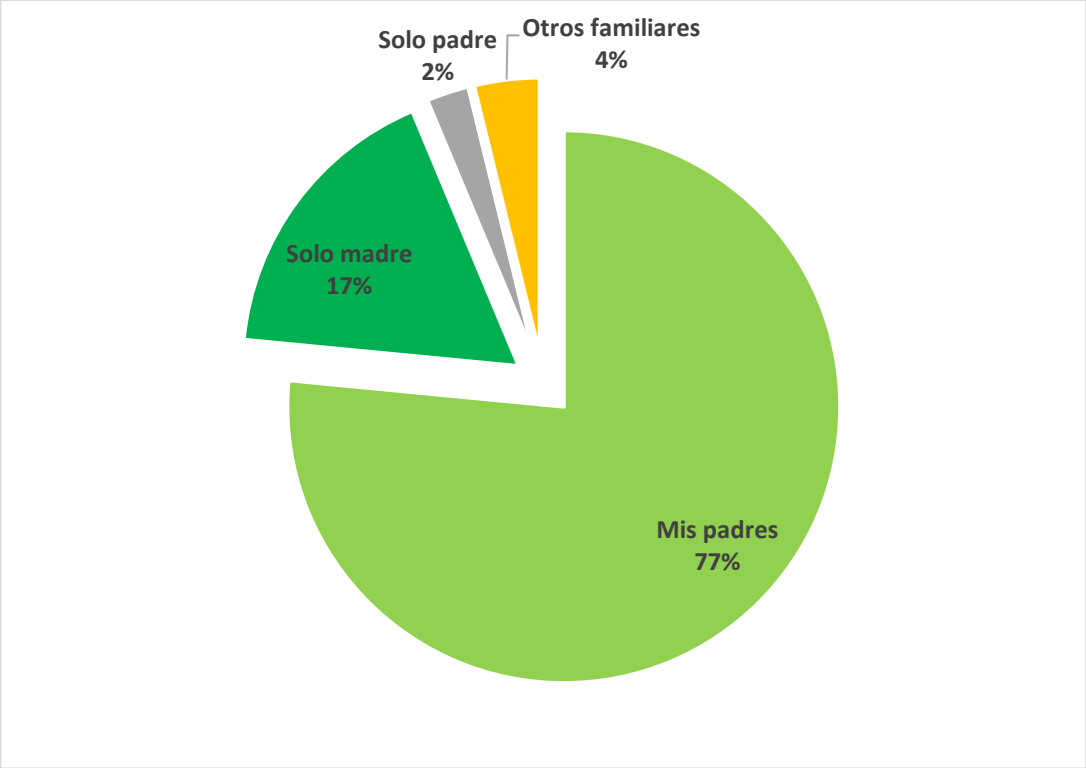


Gráfico 6.

Según la edad de los estudiantes se observa: la edad promedio de los estudiantes (14.61 ± 1.705), el 50% de los estudiantes tienen una edad menor o igual a 15 años y el otro 50% supero esta cantidad siendo la edad mínima 11 y la edad máxima de 19 años.

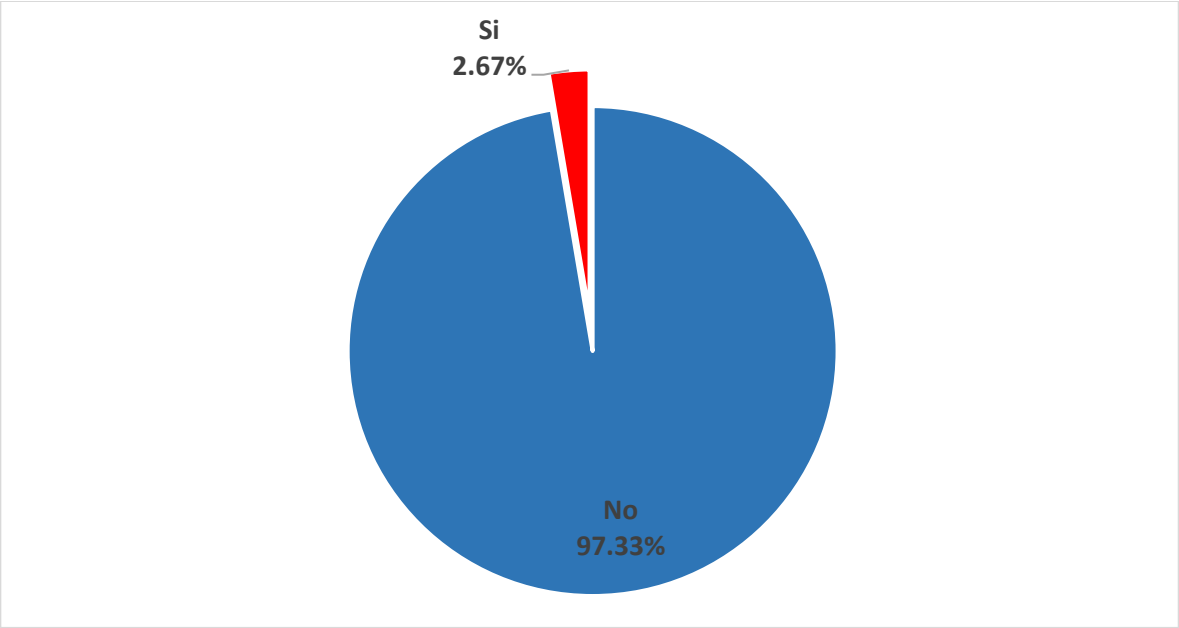
Gráfico 7. Convivencia de los estudiantes de la Institución Educativa de Lambayeque, Julio 2018.



Fuente: Datos recolectados de la Oficina de Tutoría de la I. E. – Julio 2018.
Elaboración propia

Según los resultados se observa que el 76.58% de los estudiantes convive con sus padres, el 17.18% solo con su madre, el 2.48% con su padre y el 3.82% con otros familiares.

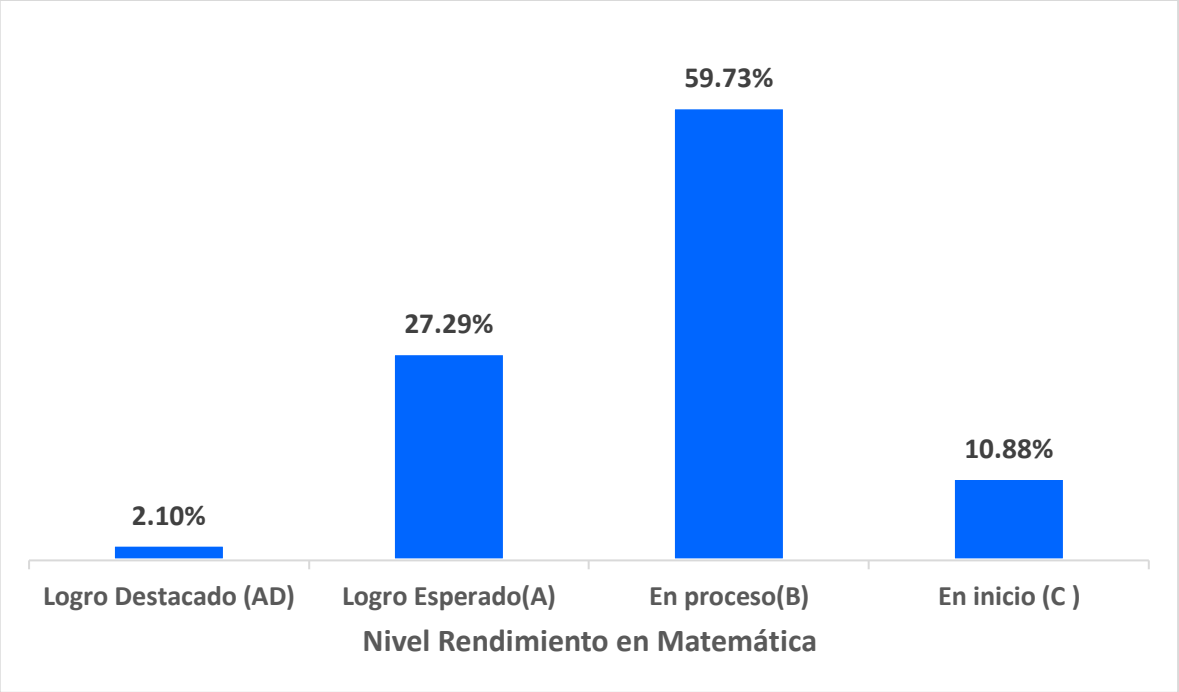
Gráfico 8. Actualmente trabaja según los estudiantes de la Institución Educativa de Lambayeque, Julio 2018.



Fuente: Datos recolectados de la Oficina de Tutoría de la I. E. – Julio 2018.
Elaboración propia

Según la variable en estudio si actualmente trabaja: los estudiantes manifestaron el 2.67% si trabajan y el 97.33% no trabajan.

Gráfico 9. Rendimiento en Matemática de los estudiantes de la Institución Educativa de Lambayeque, Julio 2018.



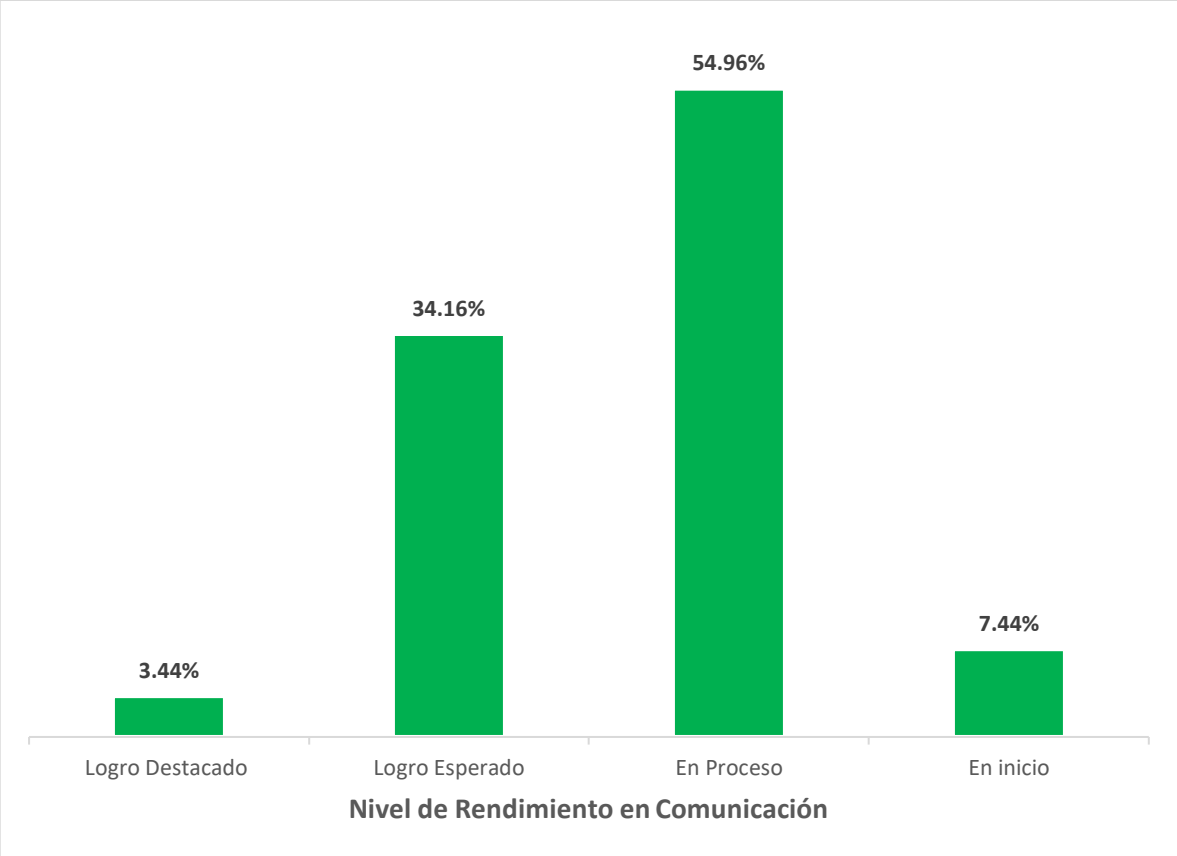
Fuente: Datos recolectados de la Oficina de Tutoría de la I. E. – Julio 2018.

Elaboración propia

El rendimiento en matemática es una las variables importantes en el modelo el cuál siendo la nota mínima de 01 y la máxima de 19 puntos, La mayoría de los estudiantes tienen una nota de 12 (22.33%) la cuál es regular.

Según el rendimiento de los estudiantes se observa: en nivel de inicio (10.88%), en proceso (59.73%); logro esperado (27.29%) y en nivel de logro destacado (2.10%).

Gráfico 10 . Rendimiento en Comunicación Integral de los estudiantes de la Institución Educativa de Lambayeque, Julio 2018.



Fuente: Datos recolectados de la Oficina de Tutoría de la I. E. – Julio 2018.
Elaboración propia

El rendimiento en el área de comunicación integral se observa que la puntuación mínima es de 7 puntos y la máxima es de 19 puntos.

Según los niveles se observa que el 7.44% de los estudiantes en el área de comunicación integral: el 7.44% nivel de inicio (0 – 10); la mayoría de los alumnos 62.40% tienen una puntuación de mayor o igual a 7 pero menor de 14.

Tabla 6. Medidas descriptivas del Rendimiento en Matemática y Comunicación Integral de los Estudiantes de una Institución Educativa de la ciudad de Lambayeque, Julio 2018.

Medidas Descriptivas	MATEMATICA	COMUNICACIÓN
Media	12.73	13.16
Mediana	12	13
Moda	12	12
Desviación estándar	2.179	2.166
Coeficiente de Asimetría	0.2862	0.3827
Coeficiente de Curtosis	1.259	0.005602
Minimo	1	7
Maximo	19	19
Percentil 25	11	12
Percentil 50	12	13
Percentil 75	14	14.75

Fuente: Datos recolectados de la Oficina de Tutoría de la I. E. – Julio 2018.
Elaboración propia

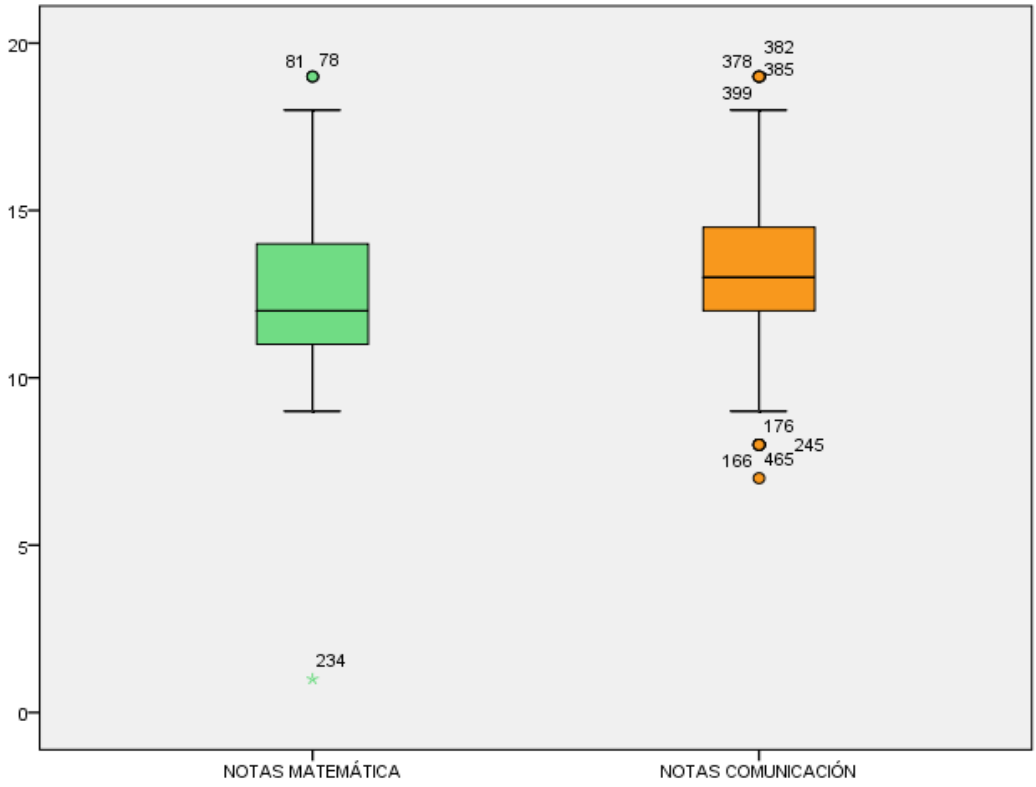


Gráfico 11.

Según la puntuación total del rendimiento en los cursos de matemática y comunicación integral se observa:

La puntuación promedio en el rendimiento en matemática es de (12.73± 2.179), el 50% de los estudiantes obtuvieran una puntuación menor o igual a 12 y el otro 50% supero esta cantidad siendo la nota mínima de 1 y la puntuación máxima de 19.

La puntuación en comunicación integral la nota promedio es (13.16 ± 2.166), el 50% de los estudiantes tuvieron una nota menor o igual a 13 y el otro 50% supera esta cantidad siendo la nota mínima de 7 y la máxima de 19.

Tabla 7. Prueba Binomial de la Variable Conducta Antisocial

Prueba binomial

Riesgo de Conducta Antisocial	Categoría	N	Prop. observada	Prop. de prueba	Significación exacta (bilateral)
Grupo 1	No	503	,96	,50	,000
Grupo 2	Si	21	,04		
Total		**524**	**1,00**		

Al realizar el análisis de las pruebas estadística aplicadas, se concluye que están asociadas las variables predictora y la variable dependiente. Realizando un análisis de correlación excluyendo a las variables que no aportan al modelo de Regresión.

3.1.2. Aplicación de la Regresión Logística

Al realizar el análisis mediante Regresión logística Binaria para identificar las variables significativas que aportan al modelo para cada uno de los casos obtuvimos:

Paso 1 . Eliminar variables que no se usarán

Se realizara el análisis identificando las variables independientes GRADO, (1ro, 4to, 5to año) de educación Básica del nivel secundaria, SEXO (Masculino , Femenino), EDAD del estudiante, TRABAJA (si actualmente el estudiante trabaja para apoyar el ingreso familiar), CONVIVENCIA (con quien vive actualmente), MATEMÁTICA (Rendimiento en Matemática), COMUNICACIÓN (Rendimiento en Comunicación). La variable Dependiente CONDUCTA (Riesgo de conducta antisocial).

Para el caso de la variable dependiente No (0), Si (1). Lo cual representa en cifras porcentuales no están en riesgo (95.99%) , Si están en riesgo (4.01%).

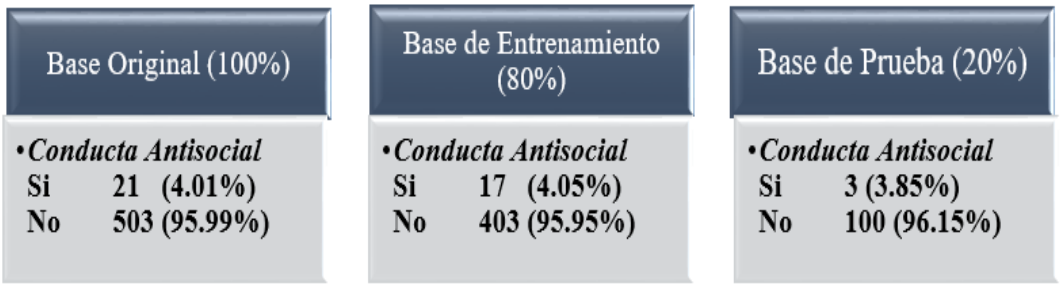
Tabla 8. **Variables finales para el modelamiento**

Variable	Tipo
Grado	Categórica
Sexo	Categórica
Edad	Numérica
Convivencia	Categórica
Trabaja	Categórica
Rendimiento en Matemática	Numérica
Rendimiento en Comunicación	Numérica
Conducta	Categórica

Fuente: Elaboración propia.

Paso 2. Selección de muestra de entrenamiento (80%) y de prueba (20%)

Para este paso, se dividió la data en una muestra de entrenamiento (training) y una muestra de prueba (testing) en una proporción de 8:2. Es decir, se tiene un 80% de datos para entrenar el modelo y un 20% de datos para probar el modelo. Para el pre procesamiento de datos se tiene la distribución que se presenta en la Gráfica 12.



Gráfica 12. Selección de muestras

Paso 3. Regresión logística con datos sin balancear

Ejecutar un modelo lógico (Tabla 13) para equilibrar las categorías sin procesamiento previo de datos es un paso importante para ignorar, porque la regresión no puede "aprender" porque las categorías tienen datos insuficientes, en este caso una minoría de categorías. (Si) =1, están en riesgo de conducta antisocial.

Se realizará el análisis identificando las variables independientes GRADO, (1ro, 4to, 5to año) de educación Básica del nivel secundaria, SEXO (Masculino, Femenino), EDAD del estudiante, TRABAJA (si actualmente el estudiante trabaja para apoyar el ingreso familiar), CONVIVENCIA (con quien vive actualmente), MATEMÁTICA (Rendimiento en Matemática), COMUNICACIÓN (Rendimiento académico en Comunicación). La variable Dependiente CONDUCTA ANTISOCIAL (Riesgo de conducta antisocial).

Para el caso de la variable dependiente No (0), Si (1). Lo cual representa en cifras porcentuales no están en riesgo (95.99%), Si están en riesgo (4.01%).

Se dividió la data en una muestra de entrenamiento (training) y una muestra de prueba (testing) en una proporción de 8:2. Es decir, se tiene un 80% de datos de entrenamiento y un 20% de datos para prueba.

Balanceando los datos con el Algoritmo SMOTE

Se realiza un submuestreo aleatorio simple en la mayoría de las categorías para equilibrar los datos, utilizando el algoritmo "SMOTE, en R-project, la proporción de los datos de la biblioteca DMwR que se muestra en la tabla, se mantiene de la siguiente manera.

Tabla 9. Balanceando los datos con el Algoritmo SMOTE

Riesgo de Conducta Antisocial	N°	%
Si	119	43.8%
No	153	56.3%
Total	**272**	**100.0%**

Fuente: Elaboración Propia

Paso 4. Aplicando la Regresión Logística a los datos balanceados con SMOTE

Tabla 10: Regresión logística con balanceo mediante SMOTE

Variable	Coeficientes	Error Estándar	Valor Z	Pr (> |z |)	Exp (B)
(Intercepto)	-5.14537	1369.79766	-0.0040	0.997003	0.005826306
GRADOCUARTO	19.79291	1369.79323	0.0140	0.988471	394412860.8
GRADOQUINTO GRADO	20.53403	1369.7933	0.0150	0.98804	827590926.5
SEXOMasculino	0.9736	0.5019	1.9400	0.052401	2.647452294
EDAD_1	-0.24347	0.23308	-1.0450	0.29622	0.783906583
TRABAJASi	19.35049	3174.56475	0.0060	0.995137	253402177.5
CONVIVENCIASolo madre	0.07692	0.46828	0.1640	0.869531	1.079951947
CONVIVENCIASolo padre	3.34083	1.11334	3.0010	0.002693 **	28.24243762
CONVIVENCIAOtros familiares	19.22703388	4239.99826	0.0051	0.996383	4.4877E-09
MATEMATICA	0.554950144	0.14383	3.8580	0.000114 ***	0.574101015
COMUNICACIÓN	0.495560129	0.12906	3.8400	0.0000123***	0.609227893

FUENTE: Elaboración propia con R-Studio.

El resultado de la predicción en la tabla de confusión (Tabla 11), muestra una cantidad de estudiantes clasificados correctamente presenta riesgo (Si), así como el número de estudiantes clasificados en forma correcta No en riesgo.

Tabla 11: Tabla Cruzada de Regresión Logística con SMOTE

Real	Predicho		Total
	No	Si	
No	80	0	80
Si	20	4	24
Total	100	4	104

FUENTE: Elaboración propia con R-Studio.

Observando la tabla del índice de clasificación (Tabla 13), se confirma que ambas clasificaciones han mejorado: una minoría, con una sensibilidad del 100%, y una mayoría

con una especificidad del 80%. De manera similar, una precisión del 80,77% es aceptable y el área bajo la curva es del 84,25%.

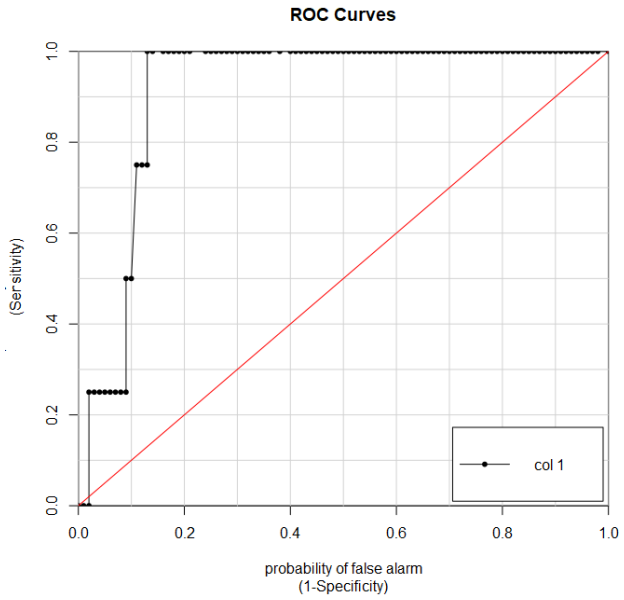
Tabla 12. Indicadores de clasificación de una Regresión Logística con SMOTE

	Regresión Logística SMOTE
Sensibilidad	1.000000
Especificidad	0.8
ROC	0.8425
Precisión/C. Global	0.8077

FUENTE: Elaboración propia con R-Studio.

Al realizar la regresión logística con SMOTE, obtiene un alto índice de sensibilidad, por tanto, predice mejor la categoría Si está en Riesgo de Conducta Antisocial. Según al resultado de la curva ROC (Gráfica), el área bajo la curva es de 84.25% clasifique correctamente a los estudiantes.

Aplicando la Regresión Logística con SMOTE tiene alto desempeño como clasificador ante el riesgo de conducta antisocial.



Gráfica 13. Áreas bajo la curva (AUC)

Fuente: Elaboración Propia.

Variables que influyen en el Riesgo de Conducta antisocial

Se determinó las siguientes variables significativas en el modelo para poder predecir el riesgo de conducta antisocial son: Grado, sexo, edad, convivencia, rendimiento en Matemática, rendimiento en comunicación

Predicción de nuevos estudiantes.

Se realizó el proceso de predicción de estudiantes en Riesgo de Conducta Antisocial con el modelo de regresión logística binaria con sobre-muestreo aplicando el algoritmo de SMOTE. La regresión es la siguiente:

y = 1/ (1 + exp(- (-5.15 – 1.9.79 (gradocuarto) + 20.53 (gradoquinto) + 0.97 (sexomasculino) -0.24 (edad) + 19.35 (trabajasi) + 0.08 (convivenciasolomadre) + 3.34(convivenciasolopadre)- 19.22 (convivenciaotros familiares) – 0.55 (matemática) – 0.50 (comunicación))

En la tabla 13 se presenta 6 conjuntos de datos de estudiantes, con sus respectivas variables predictoras. Con el fin de clasificar a cada estudiante como: Si (en riesgo) o *No (No riesgo)*

Se presentan las probabilidades, y la clasificación final.

Tabla 13: Predicción de nuevos estudiantes

ID	GRADO	SEXO	EDAD	TRABAJA	CONVIVENCIA	MATEMÁTICA	COMUNICACIÓN	CONDUCTA	CLASE.MODELO LOGÍSTICO	PROBABILIDAD
6	Primer grado	Femenino	13	No	Mis padres	11	10	No	No	3.86811E-09
8	Primer grado	Femenino	12	No	Mis padres	14	12	No	No	3.46544E-10
9	Primer grado	Masculino	12	No	Solo madre	12	12	No	No	3.00618E-09
10	Primer grado	Masculino	11	No	Mis padres	17	15	No	No	5.00757E-11
18	Primer grado	Masculino	12	No	Mis padres	14	14	No	No	3.40523E-10
30	Primer grado	Masculino	11	No	Otros familiares	13	14	No	No	2.22045E-16

Fuente: Elaboración Propia.

3.2. DISCUSIÓN

Esta investigación tuvo como como objetivo determinar un modelo predictivo aplicando la Regresión Logística para identificar los factores de riesgo y protección conducta antisocial en adolescentes de una Institución Educativa de la ciudad de Lambayeque.

A continuación, se estarán discutiendo los principales hallazgos de este estudio.

1. Para la construcción de los modelos se debe seleccionar una muestra de entrenamiento (80%) y prueba (20%) de la base de datos original. Por tanto, se dividió la data en una muestra de entrenamiento (training) y una muestra de prueba (testing) en una proporción de 8:2.

2. Cuando utilice SMOTE para la regresión logística, obtendrá un índice de sensibilidad alto, puede predecir mejor la categoría de Si está en riesgo de comportamiento antisocial. Según los resultados de la curva ROC, el área bajo la curva es del 84,25% para clasificar correctamente a los estudiantes. Ante el riesgo de comportamiento antisocial, aplicar la regresión logística junto con SMOTE como clasificador tiene un mejor desempeño.

3. Se identificando las variables significativas en el modelo para poder predecir el riesgo de conducta antisocial son: Grado, sexo, edad, convivencia, rendimiento en Matemática, rendimiento en comunicación.

4. Estudios similares en otras poblaciones como por ejemplo en jugadores de futbol el análisis de regresión muestra que la orientación a la tarea y el clima de dominación son predictores positivos de comportamiento prosocial, mientras que la autodirección y el clima de desempeño son predictores positivos de comportamiento antisocial. Además, la orientación a la tarea predice negativamente el comportamiento antisocial, mientras que la autoorientación predice negativamente el comportamiento prosocial (Kavussanu, 2006). Los modelos de regresión ayudan también a predecir comportamientos en adolescentes, de acuerdo con muchos modelos de regresión logística, los factores de

riesgo importantes para las conductas antisociales de los adolescentes incluyen las conductas problemáticas previas de los niños (es decir, agresividad y problemas de atención / irritabilidad a los 5 años) e inestabilidad marital, lo que aumenta el riesgo de conductas antisociales. Los factores perinatales, el uso de materiales maternos y los estilos de crianza son predictores relativamente pobres de la conducta antisocial (Bor et al., 2004).

Se utilizó un análisis de regresión para examinar la relación entre la perpetración del acoso entre los 13 y los 14 años y el comportamiento antisocial. Se encontraron asociaciones significativas moderadas entre la perpetración del acoso y el comportamiento antisocial posterior (Renda et al., 2011). En otro estudio el análisis de regresión jerárquica muestra que los efectos de interacción indican que las mismas variables pueden lograr funciones de riesgo y protección: la pertenencia a un grupo reducido y la satisfacción con el apoyo social promueven la continuidad del comportamiento (Bender & Lösel, 1997). Análisis de regresión sirvieron para determinar que la intimidación es un poderoso predictor de casi todos los resultados antisociales. El acoso físico es más predictivo que el verbal / indirecto. Controlar los riesgos domésticos y los problemas de externalización / internalización reducirá la escala del impacto, pero el acoso escolar sigue siendo un predictor fiable (Bender & Lösel, 2011). Un modelo de regresión logística jerárquica se puede usar en otros casos, que considera factores de confusión y puntajes de consumo de alcohol encontró que los atletas universitarios tienen más probabilidades de exhibir un comportamiento agresivo que los no atletas por ejemplo (O’Brien et al., 2012).

IV. CAPITULO IV. CONCLUSIONES

1. Los resultados obtenidos como: indicadores; clasificación global, curva ROC sensibilidad y especificidad. La regresión logística binaria aplicando la técnica de balanceo: algoritmo de SMOTE, obtiene mejores resultados en la clasificación de estudiantes en riesgo de una conducta antisocial.

2. Aplicando la Regresión Logística con SMOTE posee mejor desempeño como clasificador ante el riesgo de conducta antisocial; obtiene un alto índice de sensibilidad, es decir predice mejor la categoría de Si está en Riesgo de Conducta Antisocial. En cuanto a la curva ROC, el área bajo la curva es de 84.25% clasifique correctamente a los estudiantes.

3. Se identificó las variables significativas en el modelo para poder predecir el riesgo de conducta antisocial son: Grado, sexo, edad, convivencia, rendimiento en Matemática, rendimiento en comunicación.

V. CAPITULO V: RECOMENDACIONES

1. Recomendar a otros investigadores que estén interesados en este tema, debido a su utilidad en la predicción de conductas de riesgo, con otros tipos de modelos.

2.- Implementación de un sistema para el registro de las fichas de tutoría de los estudiantes de la institución educativa, hacer seguimiento de las incidencias registradas.

3. Realizar convenios con instituciones académicas para implementación de talleres para estudiantes (habilidades sociales, control emocional, razonamiento moral, resilencia, planificación de vida, y respeto a la diversidad étnica y sexual); para los padres de familia (estilos de crianza).

BIBLIOGRAFÍA REFERENCIADA.

1. Acquaviva, E., Ellul, P., & Benarous, X. (2018). Variations in Pathways Into and Out of Antisocial Behavior From the Perspective of Developmental Psychopathology. *Understanding Uniqueness and Diversity in Child and Adolescent Mental Health*, 3–23. https://doi.org/10.1016/B978-0-12-815310-9.00001-0
2. Bender, D., & Lösel, F. (1997). Protective and risk effects of peer relations and social support on antisocial behaviour in adolescents from multi-problem milieus. *Journal of Adolescence*, *20*(6), 661–678. https://doi.org/10.1006/JADO.1997.0118
3. Bender, D., & Lösel, F. (2011). Bullying at school as a predictor of delinquency, violence and other anti-social behaviour in adulthood. *Criminal Behaviour and Mental Health*, *21*(2), 99–106. https://doi.org/10.1002/CBM.799
4. Boggio, G. (1997). Modelo de regresión logística aplicado a un estudio sobre enfermedad de Chagas. *Cadernos de Saúde Pública*, *13*(3), 389–396. https://doi.org/10.1590/S0102-311X1997000300014
5. Bor, W., McGee, T. R., & Fagan, A. A. (2004). Early risk factors for adolescent antisocial behaviour: An Australian longitudinal study. *Australian and New Zealand Journal of Psychiatry*, *38*(5), 365–372. https://doi.org/10.1111/j.1440-1614.2004.01365.x
6. Calduch Cervera, R. (2012). *Métodos y técnicas de investigación en relaciones internacionales*.
7. Calkins, S. D., & Keane, S. P. (2009). Developmental origins of early antisocial behavior. *Development and Psychopathology*, *21*(4), 1095–1109. https://doi.org/10.1017/S095457940999006X
8. Cerda, J., & Cifuentes, L. (2012). Uso de curvas ROC en investigación clínica: Aspectos teórico-prácticos. *Revista chilena de infectología*, *29*(2), 138–141. https://doi.org/10.4067/S0716-10182012000200003
9. Chávez-Bosquez, O., Torres-Vásquez, M., Hernández-Torruco, J., & Hernández-Ocaña, B. (2020). Impacto de los algoritmos de sobremuestreo en la clasificación de subtipos principales del síndrome de guillain-barré. *Ingenius*, *25*, 20–31. https://doi.org/10.17163/INGS.N25.2021.02

10. Coelho, I. C., Neves, A. C., & Caridade, S. (2020). Risk factors for antisocial behavior in children: Comparison between boys and girls. *Estudos de Psicologia (Campinas)*, *37*(1), 1. https://doi.org/10.1590/1982-0275202037e190027
11. Criss, M. M., Smith, A. M., Morris, A. S., Liu, C., & Hubbard, R. L. (2017). Parents and peers as protective factors among adolescents exposed to neighborhood risk. *Journal of Applied Developmental Psychology*, *53*, 127–138. https://doi.org/10.1016/J.APPDEV.2017.10.004
12. Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., & Osher, D. (2020). Implications for educational practice of the science of learning and development. *Applied Developmental Science*, *24*(2), 97–140. https://doi.org/10.1080/10888691.2018.1537791
13. de los Ángeles Páramo, M. (2011). Risk and protective factors in adolescence: Content analysis through discussion groups. *Terapia Psicologica*, *29*(1), 85–95. https://doi.org/10.4067/s0718-48082011000100009
14. Deković, M., Janssens, J. M. A. M., & Van As, N. M. C. (2003). Family Predictors of Antisocial Behavior in Adolescence. *Family Process*, *42*(2), 223–235. https://doi.org/10.1111/J.1545-5300.2003.42203.X
15. Díaz Rosas, F. (1995). La predicción del rendimiento académico en la universidad: un ejemplo de aplicación de la regresión múltiple. *Enseñanza & Teaching: Revista interuniversitaria de didáctica*, *13*(13), 43–62.
16. Domínguez Alonso, E., & Aldana Padilla, D. (2001). Regresión logística: Un ejemplo de su uso en Endocrinología. *Revista Cubana de Endocrinología*, *12*(1), 0–0.
17. Farrington, D. P. (2005). Childhood origins of antisocial behavior. *Clinical Psychology & Psychotherapy*, *12*(3), 177–190. https://doi.org/10.1002/CPP.448
18. Ferre Jaén, M. E. (2019). *FEIR 45: Regresión logística*. Apuntes del curso FEIR3.
19. Frick, P. J., & Viding, E. (2009). Antisocial behavior from a developmental psychopathology perspective. *Development and Psychopathology*, *21*(4), 1111–1131. https://doi.org/10.1017/S0954579409990071
20. Gard, A. M., Dotterer, H. L., & Hyde, L. W. (2019). Genetic influences on antisocial behavior: recent advances and future directions. *Current Opinion in Psychology*, *27*, 46–55. https://doi.org/10.1016/J.COPSYC.2018.07.013

21. Gershoff, E. T. (2002). Corporal punishment by parents and associated child behaviors and experiences: a meta-analytic and theoretical review. *Psychological bulletin*, *128*(4), 539–579. https://doi.org/10.1037/0033-2909.128.4.539
22. Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2000). *Applied logistic regression*. Wiley New York.
23. Huesmann, L. R., & Taylor, L. D. (2006). The role of media violence in violent behavior. En *Annual Review of Public Health* (Vol. 27, pp. 393–415). Annual Reviews. https://doi.org/10.1146/annurev.publhealth.26.021304.144640
24. Jaume, M. J., & Morar-Catalá, R. (2001). Análisis de Regresión Múltiple. En *Estadística Informática: casos y ejemplos con el SPSS* (pp. 3–17).
25. Justicia, F., Benítez, J. L., Pichardo, M. C., Fernández, E., García, T., & Fernández, M. (2006). Towards a new explicative model of antisocial behaviour. *Electronic Journal of Research in Educational Psychology*, *4*(9), 131–150. https://doi.org/10.25115/ejrep.v4i9.1187
26. Kavussanu, M. (2006). Motivational predictors of prosocial and antisocial behaviour in football. *Journal of Sports Sciences*, *24*(6), 575–588. https://doi.org/10.1080/02640410500190825
27. López-Roldán, P., & Fachelli, S. (2018). Metodologia de la Investigaciin Social Cuantitativa (2015). En *Revista de Educacion y Derecho* (Número 17). https://doi.org/10.1344/reyd2018.17.13
28. Lösel, F., & Bender, D. (2017). Protective Factors Against Crime and Violence in Adolesence. *The Wiley Handbook of Violence and Aggression*, 1–15. https://doi.org/10.1002/9781119057574.WHBVA025
29. Mahoney, J. L., & Stattin, H. (2000). Leisure activities and adolescent antisocial behavior: The role of structure and social context. *Journal of Adolescence*, *23*(2), 113–127. https://doi.org/10.1006/JADO.2000.0302
30. Murray, J., Shenderovich, Y., Gardner, F., Mikton, C., Derzon, J. H., Liu, J., & Eisner, M. (2018). Risk factors for antisocial behavior in low- and middle-income countries: A systematic review of longitudinal studies. En *Crime and Justice* (Vol. 47, Número 1, pp. 255–364). University of Chicago Press. https://doi.org/10.1086/696590
31. O'Brien, K. S., Kolt, G. S., Martens, M. P., Ruffman, T., Miller, P. G., & Lynott, D.

(2012). Alcohol-related aggression and antisocial behaviour in sportspeople/athletes. *Journal of Science and Medicine in Sport*, *15*(4), 292–297. https://doi.org/10.1016/J.JSAMS.2011.10.008

32. Office of the Surgeon General. (2001). Chapter 4 -- Risk Factors for Youth Violence. En *Youth Violence: A Report of the Surgeon General* (Número Md, p. 233). Office of the Surgeon General (US).
33. Olmedo Hernández, I. J., Denis Rodríguez, E., Barradas Alarcón, M. E., Villegas Domínguez, J. E., & Denis Rodríguez, P. B. (2019). Agresividad y conducta antisocial en individuos con dependencia al teléfono móvil: un posible factor criminogénico. *Horizonte Médico (Lima)*, *19*(3), 12–19. https://doi.org/10.24265/HORIZMED.2019.V19N3.03
34. Pedrosa, A. L., Bitencourt, L., Fróes, A. C. F., Cazumbá, M. L. B., Campos, R. G. B., de Brito, S. B. C. S., & Simões e Silva, A. C. (2020). Emotional, Behavioral, and Psychological Impact of the COVID-19 Pandemic. *Frontiers in Psychology*, *11*, 2635. https://doi.org/10.3389/FPSYG.2020.566212/BIBTEX
35. Peña Fernández, M. E. de la, & Peña Fernández, M. E. de la. (2005). *Conducta antisocial en adolescentes: factores de riesgo y de protección*. Universidad Complutense de Madrid, Servicio de Publicaciones.
36. Renda, J., Vassallo, S., & Edwards, B. (2011). Bullying in early adolescence and its association with anti-social behaviour, criminality and violence 6 and 10 years later. *Criminal Behaviour and Mental Health*, *21*(2), 117–127. https://doi.org/10.1002/CBM.805
37. Reyes Rocabado, J., Escobar Flores, C., Duarte Vargas, J., & Ramírez Peradotto, P. (2007). UNA APLICACIÓN DEL MODELO DE REGRESIÓN LOGÍSTICA EN LA PREDICCIÓN DEL RENDIMIENTO ESTUDIANTIL. *Estudios pedagógicos (Valdivia)*, *33*(2), 101–120. https://doi.org/10.4067/S0718-07052007000200006
38. Rivera, R., & Cahuana Cuentas, M. (2016). Influencia de la familia sobre las conductas antisociales en adolescentes de Arequipa-Perú. *Actualidades en Psicología*, *30*(120), 85. https://doi.org/10.15517/ap.v30i120.18814
39. Sugai, G., Horner, R. H., Dunlap, G., Hieneman, M., Lewis, T. J., Nelson, C. M., Scott, T., Liaupsin, C., Sailor, W., Turnbull, A. P., Turnbull, H. R., Wickham, D., Wilcox,

B., & Ruef, M. (2000). Applying Positive Behavior Support and Functional Behavioral Assessment in Schools. *Journal of Positive Behavior Interventions*, *2*(3), 131–143. https://doi.org/10.1177/109830070000200302

40. Uriarte Arciniega, J. (2005). La resiliencia. Una nueva perspectiva en psicopatología del desarrollo. *Revista de psicodidáctica*, *10*(2), 61–79.
41. van Hazebroek, B. C. M., Wermink, H., van Domburgh, L., de Keijser, J. W., Hoeve, M., & Popma, A. (2019). Biosocial studies of antisocial behavior: A systematic review of interactions between peri/prenatal complications, psychophysiological parameters, and social risk factors. En *Aggression and Violent Behavior* (Vol. 47, pp. 169–188). Pergamon. https://doi.org/10.1016/j.avb.2019.02.016
42. Vega-Vilca, J. C., & Guzman, J. (2011). Regresion pls y pca como solucion al problema de multicolinealidad en regresion multiple. *Revista de Matematica: Teoria y Aplicaciones*, *18*(1), 9–20.
43. Walker, H. M., Kavanagh, K., Stiller, B., Golly, A., Severson, H. H., & Feil, E. G. (1998). First Step to Success: An Early Intervention Approach for Preventing School Antisocial Behavior. *Journal of Emotional and Behavioral Disorders*, *6*(2), 66–80. https://doi.org/10.1177/106342669800600201

ANEXOS

ANEXO 1. Regresión logística en R – Studio.

```
str(datos.m)
'data.frame':	524 obs. of  8 variables:
 $ GRADO       : Factor w/ 3 levels "PRIMER GRADO",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ SEXO        : Factor w/ 2 levels "Femenino","Masculino": 2 2 2 2 2 1 2 1 2 2 ...
 $ EDAD_1      : num  12 12 12 12 11 13 12 12 12 11 ...
 $ TRABAJA     : Factor w/ 2 levels "No","Si": 1 1 1 1 1 1 1 1 1 1 ...
 $ CONVIVENCIA : Factor w/ 4 levels "Mis padres","Solo madre",..: 1 1 1 1 1 1 1 1 2 1 ...
 $ MATEMATICA  : num  13 17 13 14 13 11 17 14 12 17 ...
 $ COMUNICACIÓN: num  10 11 11 12 11 10 12 12 12 15 ...
 $ CONDUCTA    : Factor w/ 2 levels "No","Si": 1 1 1 1 1 1 1 1 1 1 ...
>
> # 1. Eliminar variables que no se usarán
> datos.m <- datos.m[,c(2,6,8:10,17:19)]
Error in `[.data.frame`(datos.m, , c(2, 6, 8:10, 17:19)) :
  undefined columns selected
> str(datos.m)
'data.frame':	524 obs. of  8 variables:
 $ GRADO       : Factor w/ 3 levels "PRIMER GRADO",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ SEXO        : Factor w/ 2 levels "Femenino","Masculino": 2 2 2 2 2 1 2 1 2 2 ...
 $ EDAD_1      : num  12 12 12 12 11 13 12 12 12 11 ...
 $ TRABAJA     : Factor w/ 2 levels "No","Si": 1 1 1 1 1 1 1 1 1 1 ...
 $ CONVIVENCIA : Factor w/ 4 levels "Mis padres","Solo madre",..: 1 1 1 1 1 1 1 1 2 1 ...
 $ MATEMATICA  : num  13 17 13 14 13 11 17 14 12 17 ...
 $ COMUNICACIÓN: num  10 11 11 12 11 10 12 12 12 15 ...
 $ CONDUCTA    : Factor w/ 2 levels "No","Si": 1 1 1 1 1 1 1 1 1 1 ...
>
> contrasts(datos.m$CONDUCTA)
   Si
No  0
Si  1
>
> round(prop.table(table(datos.m$CONDUCTA))*100,2)

   No    Si
95.99  4.01
>
```

```
> # 2. Selección de muestra de entrenamiento (80%) y de prueba (20%)
>
> library(caret)
> set.seed(123)
>
> index   <- createDataPartition(datos.m$CONDUCTA, p=0.8, list=FALSE)
> imbal_train    <- datos.m[ index, ]
> imbal_testing  <- datos.m[-index, ]
>
> # Verificando que se mantenga la proporción original
> # Data original
> addmargins(table(datos.m$CONDUCTA))

 No  Si Sum
503  21 524
> round(prop.table(table(datos.m$CONDUCTA))*100,2)

   No    Si
95.99  4.01
>
> # Data training
> addmargins(table(imbal_train$CONDUCTA))

 No  Si Sum
403  17 420
> round(prop.table(table(imbal_train$CONDUCTA))*100,2)

   No    Si
95.95  4.05
>
> # Data testing
> addmargins(table(imbal_testing$CONDUCTA))

 No  Si Sum
100   4 104
> round(prop.table(table(imbal_testing$CONDUCTA))*100,2)

   No    Si
96.15  3.85
>
```

```
> #-----------------------------------------------------------------
> # 3. Balanceando los datos con el algoritmo Smote
> library(DMwR)

> set.seed(123)
>
> smote_train <- SMOTE(CONDUCTA ~ .,
+                      data=imbal_train,
+                      perc.over = 600,
+                      perc.under =150)
>
> addmargins(table(smote_train$CONDUCTA))

 No  Si Sum
153 119 272
> addmargins(table(smote_train$CONDUCTA) )

 No  Si Sum
153 119 272
>
> #-----------------------------------------------------------------
> # 4. Aplicando la Regresión Logística a los datos balanceados
> #    con SMOTE
> options(scipen=999)
> modelo_logistico <- glm(CONDUCTA ~ . ,
+                         family=binomial,
+                         data=smote_train)
>
> summary(modelo_logistico)

Call:
glm(formula = CONDUCTA ~ ., family = binomial, data = smote_train)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.15575  -0.54209  -0.00004   0.55860   2.21211


Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)               -5.14537 1369.79766  -0.004 0.997003
GRADOCUARTO GRADO         19.79291 1369.79323   0.014 0.988471
GRADOQUINTO GRADO         20.53403 1369.79330   0.015 0.988040
SEXOMasculino              0.97360    0.50190   1.940 0.052401 .
EDAD_1                    -0.24347    0.23308  -1.045 0.296220
TRABAJASi                 19.35049 3174.56475   0.006 0.995137
CONVIVENCIASolo madre      0.07692    0.46828   0.164 0.869531
CONVIVENCIASolo padre      3.34083    1.11334   3.001 0.002693 **
CONVIVENCIAOtros familiares -19.22193 4239.99826  -0.005 0.996383
MATEMATICA                -0.55495    0.14383  -3.858 0.000114 ***
COMUNICACIÓN              -0.49556    0.12906  -3.840 0.000123 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 372.81  on 271  degrees of freedom
Residual deviance: 196.46  on 261  degrees of freedom
AIC: 218.46

Number of Fisher Scoring iterations: 18

> coef(modelo_logistico)
               (Intercept)          GRADOCUARTO GRADO          GRADOQUINTO GRA
DO
               -5.14537205                19.79290879                 20.534029
54
             SEXOMasculino                     EDAD_1                  TRABAJA
Si
                0.97359778                -0.24346542                 19.350488
42
```

```
      CONVIVENCIASolo madre       CONVIVENCIASolo padre CONVIVENCIAOtros familiar
es
                 0.07691655                  3.34082573                -19.221926
34
                 MATEMATICA                COMUNICACIÓN
                -0.55494991                 -0.49556287
>
>
> # Cociente de ventajas (Odd Ratio)
> exp(coef(modelo_logistico))
                (Intercept)           GRADOCUARTO GRADO           GRADOQUINTO GRA
DO
          0.005826306287898   394412860.834417343139648   827590926.4678016901016
24
              SEXOMasculino                      EDAD_1                   TRABAJA
Si
          2.647452294303134           0.783906583003690   253402177.5275387167930
60
      CONVIVENCIASolo madre       CONVIVENCIASolo padre CONVIVENCIAOtros familiar
es
          1.079951947187530          28.242437624014762           0.0000000044876
96
                 MATEMATICA                COMUNICACIÓN
          0.574101014762303           0.609227892511972
>
> # Coeficientes y Odd-Ratio del modelo logístico
> cbind(Coeficientes=modelo_logistico$coef,ExpB=exp(modelo_logistico$coef))
                           Coeficientes                     ExpB
(Intercept)                 -5.14537205         0.005826306287898
GRADOCUARTO GRADO           19.79290879 394412860.834417343139648
GRADOQUINTO GRADO           20.53402954 827590926.467801690101624
SEXOMasculino                0.97359778         2.647452294303134
EDAD_1                      -0.24346542         0.783906583003690
TRABAJASi                   19.35048842 253402177.527538716793060
CONVIVENCIASolo madre        0.07691655         1.079951947187530
CONVIVENCIASolo padre        3.34082573        28.242437624014762
CONVIVENCIAOtros familiares -19.22192634         0.000000004487696
MATEMATICA                  -0.55494991         0.574101014762303
COMUNICACIÓN                -0.49556287         0.609227892511972
>
```

```
> # Cociente de ventajas e Intervalo de Confianza al 95%
> library(MASS)
> exp(cbind(OR = coef(modelo_logistico),confint.default(modelo_logistico)))
                                          OR     2.5 %      97.5 %
(Intercept)                     0.005826306287898 0.0000000         Inf
GRADOCUARTO GRADO             394412860.834417343139648 0.0000000         Inf
GRADOQUINTO GRADO             827590926.467801690101624 0.0000000         Inf
SEXOMasculino                   2.647452294303134 0.9899437   7.0802040
EDAD_1                          0.783906583003690 0.4964431   1.2378248
TRABAJASi                     253402177.527538716793060 0.0000000         Inf
CONVIVENCIASolo madre           1.079951947187530 0.4313253   2.7039831
CONVIVENCIASolo padre          28.242437624014762 3.1858774 250.3659659
CONVIVENCIAOtros familiares     0.000000004487696 0.0000000         Inf
MATEMATICA                      0.574101014762303 0.4330735   0.7610532
COMUNICACIÓN                    0.609227892511972 0.4730718   0.7845715
>
> #-----------------------------------------------------------------
> # 5. Predicción del modelo logistico en la data testing (data que
> #    no se balancea)
> proba.modelo_logistico <- predict(modelo_logistico,
+                                   imbal_testing,
+                                   type="response")
> head(proba.modelo_logistico)
                        6                        8                        9
0.000000003868108125495 1 0.000000000346544328918 2 0.000000003006178337211 0
                       10                       18                       30
0.000000000050075685957 3 0.000000000340523035728 5 0.000000000000000222044 6
>
> clase.modelo_logistico <- ifelse(proba.modelo_logistico >= 0.5, 1, 0)
> clase.modelo_logistico <- factor(clase.modelo_logistico,
+                                  levels=c(0,1),
+                              labels=c("No","Si"))
```

```
> head(cbind(imbal_testing,clase.modelo_logistico,proba.modelo_logistico))
          GRADO      SEXO EDAD_1 TRABAJA      CONVIVENCIA MATEMATICA COMUNICACIÓN
CONDUCTA
6  PRIMER GRADO  Femenino     13      No       Mis padres         11           10
No
8  PRIMER GRADO  Femenino     12      No       Mis padres         14           12
No
9  PRIMER GRADO Masculino     12      No       Solo madre         12           12
No
10 PRIMER GRADO Masculino     11      No       Mis padres         17           15
No
18 PRIMER GRADO Masculino     12      No       Mis padres         14           14
No
30 PRIMER GRADO Masculino     11      No Otros familiares         13           14
No
   clase.modelo_logistico   proba.modelo_logistico
6                      No 0.0000000038681081254951
8                      No 0.0000000003465443289182
9                      No 0.0000000030061783372110
10                     No 0.0000000000500756859573
18                     No 0.0000000003405230357285
30                     No 0.0000000000000002220446
> datos.juntos <- cbind(imbal_testing,clase.modelo_logistico,proba.modelo_logisti
co)
>

> #---------------------------------------------
> # 6. Indicadores del modelo (obtenidos en la data testing)
>
> # 6.1.Matriz de confusión
> caret::confusionMatrix(clase.modelo_logistico,
+                        imbal_testing$CONDUCTA,
+                        positive="Si")
Confusion Matrix and Statistics

          Reference
Prediction No Si
        No 80  0
        Si 20  4

               Accuracy : 0.8077
                 95% CI : (0.7187, 0.8784)
    No Information Rate : 0.9615
    P-Value [Acc > NIR] : 1

                  Kappa : 0.2353
 Mcnemar's Test P-Value : 0.00002152

            Sensitivity : 1.00000
            Specificity : 0.80000
         Pos Pred Value : 0.16667
         Neg Pred Value : 1.00000
             Prevalence : 0.03846
         Detection Rate : 0.03846
   Detection Prevalence : 0.23077
      Balanced Accuracy : 0.90000

       'Positive' Class : Si

>
> result <- caret::confusionMatrix(clase.modelo_logistico,
+                                  imbal_testing$CONDUCTA,
+                                  positive="Si")
>
> result$byClass["Sensitivity"]
Sensitivity
          1
> result$byClass["Specificity"]
Specificity
        0.8
> result$overall["Accuracy"]
 Accuracy
0.8076923
```

```
>
> # 6.2. Para obtener el área bajo la curva (ROC)
> library(caTools)
> colAUC(proba.modelo_logistico,
+        imbal_testing$CONDUCTA,
+        plotROC = TRUE)
             [,1]
No vs. Si 0.8425
> abline(0, 1,col="red")
>
>
> write.csv(datos.juntos,"datos_testing_predicciones.csv")
```

ANEXO 2. Fotografías de los Programas que realiza la Policía Nacional de la Oficina de Participación Ciudadana, con los estudiantes de la Institución Educativa.

Fotografía 1. Trotes Motivacionales a cargo de los Miembros de la Policía Nacional.

Fuente: Oficina de Participación Ciudadana de la Ciudad de Lambayeque

Fotografía 2. Trotes Motivacionales en las calles principales de la Ciudad de Lambayeque.

Fuente: Oficina de Participación Ciudadana de la Ciudad de Lambayeque

ANEXO 3

Informes de similitud del software Turnitin

Modelo de regresión logística en la predicción de la conducta antisocial en adolescentes de una Institución Educativa de la ciudad de Lambayeque. 2017

INFORME DE ORIGINALIDAD

19%	19%	2%	%
INDICE DE SIMILITUD	FUENTES DE INTERNET	PUBLICACIONES	TRABAJOS DEL ESTUDIANTE

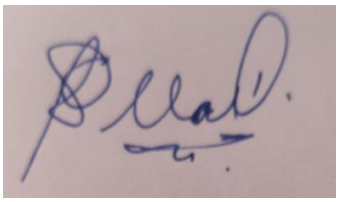
FUENTES PRIMARIAS

1	repositorio.lamolina.edu.pe Fuente de Internet	5%
2	core.ac.uk Fuente de Internet	3%
3	repositorio.unprg.edu.pe Fuente de Internet	2%
4	ddd.uab.cat Fuente de Internet	1%
5	es.slideshare.net Fuente de Internet	1%
6	docplayer.es Fuente de Internet	1%
7	repositorio.ucv.edu.pe Fuente de Internet	1%
8	fieecs.uni.edu.pe Fuente de Internet	1%

9	www.munilambayeque.gob.pe Fuente de Internet	1 %
10	1library.co Fuente de Internet	<1 %
11	eprints.ucm.es Fuente de Internet	<1 %
12	repositorio.unsm.edu.pe Fuente de Internet	<1 %
13	ichi.pro Fuente de Internet	<1 %
14	archive.org Fuente de Internet	<1 %
15	myslide.es Fuente de Internet	<1 %
16	revista-apunts.com Fuente de Internet	<1 %
17	sibu.unab.edu.co Fuente de Internet	<1 %
18	www.scribd.com Fuente de Internet	<1 %
19	difercigo.blogspot.com Fuente de Internet	<1 %
20	idoc.pub Fuente de Internet	<1 %

21	dehesa.unex.es Fuente de Internet	<1 %
22	repositorio.unfv.edu.pe Fuente de Internet	<1 %
23	repositorio.unh.edu.pe Fuente de Internet	<1 %
24	aidipe2017.aidipe.org Fuente de Internet	<1 %
25	novascientia.delasalle.edu.mx Fuente de Internet	<1 %
26	www.grafiati.com Fuente de Internet	<1 %

Excluir citas	Activo	Excluir coincidencias	< 15 words
Excluir bibliografía	Activo		



Dr. SANTIAGO BOBADILLA OCAÑA
Asesor

Recibo digital TURNITIN

turnitin

Recibo digital

Este recibo confirma quesu trabajo ha sido recibido por Turnitin. A continuación podrá ver la información del recibo con respecto a su entrega.

La primera página de tus entregas se muestra abajo.

Autor de la entrega:	Sandra Cecilia Loayza Chumacero
Título del ejercicio:	Investigaciones
Título de la entrega:	TESIS
Nombre del archivo:	INFORME_DE_TESIS_FINAL_LOAIZA.pdf
Tamaño del archivo:	1.95M
Total páginas:	69
Total de palabras:	11,104
Total de caracteres:	63,257
Fecha de entrega:	01-feb.-2023 02:00p. m. (UTC-0500)
Identificador de la entre...	2004227133

UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO

FACULTAD DE CIENCIAS HUMANOS SOCIALES Y EDUCACIÓN

UNIDAD DE POSGRADO

PROGRAMA DE MAESTRÍA EN CIENCIAS DE LA EDUCACIÓN

TESIS

Lambayeque-Perú

2023

Dr. SANTIAGO BOBADILLA OCAÑA
Asesor

Derechos de autor 2023 Turnitin. Todos los derechos reservados.

ACTA DE SUSTENTACIÓN

UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO
FACULTAD DE CIENCIAS HISTÓRICO SOCIALES Y EDUCACIÓN
UNIDAD DE INVESTIGACIÓN

ACTA DE SUSTENTACIÓN DE TESIS

N° 843-VIRTUAL

Siendo las 09:00 horas, del día martes 07 de diciembre de 2021; se reunieron vía online mediante la plataforma virtual Google Meet: https://meet.google.com/fwi-zhdw-nqu, los miembros del jurado designados mediante Resolución Nº 371-2010-UP-D-FACHSE, de fecha 21 de febrero de 2017, integrado por:

Presidente	: Dr. Jorge Isaac Castro Kikuchi.
Secretario	: Dra. María Del Pilar Fernández Celis.
Vocal	: Dra. Raquel Yovana Tello Flores.
Asesor Metodológico	: Dr. Santiago Bobadilla Ocaña.

La finalidad es evaluar la Tesis titulada: ***"MODELO DE REGRESIÓN LOGÍSTICA EN LA PREDICCIÓN DE LA CONDUCTA ANTISOCIAL EN ADOLESCENTES DE UNA INSTITUCIÓN EDUCATIVA DE LA CIUDAD DE LAMBAYEQUE 2017"***; presentada por la tesista LOAIZA CHUMACERO SANDRA CECILIA, para obtener el Grado Académico de Maestra en Ciencia de la Educación mención de Psicopedagogía Cognitiva. Producido y concluido el acto de sustentación, de conformidad con los artículos 131 al 140 del Reglamento General del Vicerrectorado de Investigación (aprobado con Resolución N° 018-2020-CU de fecha 10 de febrero del 2020); los miembros del jurado procedieron a la evaluación respectiva, haciendo las preguntas, observaciones y recomendaciones a la sustentante, quien procedió a dar respuesta a las interrogantes planteadas.

Con la deliberación correspondiente por parte del jurado, se procedió a la calificación de la Tesis, obteniendo un calificativo de (16) (DIECISEIS) en la escala vigesimal, que equivale a la mención de BUENO

Siendo las diez horas y siete minutos del mismo día, se dio por concluido el acto académico online, con la lectura del acta y la firma de los miembros del jurado.

Dr. Jorge Isaac Castro Kikuchi	**Dra. María Del Pilar Fernández Celis**	**Dra. Raquel Yovana Tello Flores**
PRESIDENTE	SECRETARIA	VOCAL

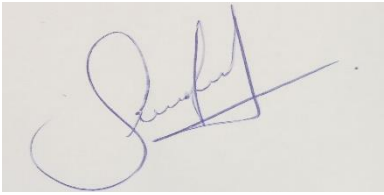
OBSERVACIONES:

El presente acto académico se sustenta en los artículos del 39 al 41 del Reglamento de Grados y Títulos de la Universidad Nacional Pedro Ruiz Gallo (aprobado con Resolución N° 270-2019-CU de fecha 4 de setiembre del 2019); la Resolución N° 407-2020-R de fecha 12 de mayo del 2020 que ratifica la Resolución N° 004-2020-VIRTUAL-VRINV del 07 de mayo del 2020 que aprueba la tramitación virtualizada para la presentación, aprobación de los proyectos de los trabajos de Investigación y de sus Informes de Investigación en cada Unidad de Investigación de las Facultades y Escuela de Posgrado; la Resolución N° 0372-2020-V-D-NG-FACHSE de fecha 21 de mayo del 2020 y su modificatoria Resolución N° 0380-2020-V-D-NG-FACHSE del 27 de mayo del 2020 que aprueba el INSTRUCTIVO PARA LA SUSTENTACIÓN DE TRABAJOS DE INVESTIGACIÓN Y TESIS VIRTUALES.

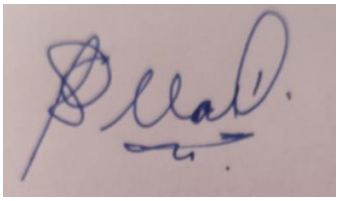
DECLARACIÓN JURADA DE ORIGINALIDAD

Yo, Sandra Cecilia Loaiza Chumacero investigador principal, y Dr. Santiago Bobadilla Ocaña asesor del trabajo de investigación "**Modelo de regresión logística en la predicción de la conducta antisocial en adolescentes de una Institución Educativa de la ciudad de Lambayeque. 2017**" Declaramos solemnemente que no hay plagio o información falsa en este trabajo. Si los hechos son contrarios, soy responsable de cancelar este informe y los trámites administrativos que puedan ocurrir. Esto puede resultar en la cancelación de títulos o grados otorgados como resultado de este informe.

Lambayeque, 09 Setiembre de 2021



SANDRA CECILIA LOAIZA CHUMACERO
Maestra



Dr. SANTIAGO BOBADILLA OCAÑA
Asesor