



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

Predicción del rendimiento académico por medio de técnicas de inteligencia artificial de los estudiantes de bachillerato en las Unidades Educativas fiscales del distrito educativo 05D06 (Salcedo - Ecuador)

KLEVER ANTONIO JIJÓN BAUTISTA

Trabajo de Titulación modalidad: Proyectos de Investigación y Desarrollo, presentado ante el Instituto de Postgrado y Educación Continua de la ESPOCH, como requisito parcial para la obtención del grado de:

**MAGÍSTER EN MATEMÁTICA, MENCIÓN MODELACIÓN Y
DOCENCIA**

Riobamba - Ecuador

Julio 2023

DECLARACIÓN DE AUTENTICIDAD

Yo, Klever Antonio Jijón Bautista, declaro que el presente proyecto de investigación es de mi autoría y que los resultados de este son auténticos y originales. Los textos constantes en el documento que provienen de otras fuentes están debidamente citados y referenciados.

Como autor, asumo la responsabilidad legal y académica de los contenidos de este Trabajo de Titulación de Maestría.



JIJÓN BAUTISTA KLEVER ANTONIO

No. Cédula 050313287-0

©2023, klever Antonio Jijón Bautista.

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento, siempre y cuando se reconozca el Derecho de Autor.



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

EL TRIBUNAL DEL TRABAJO DE TITULACIÓN CERTIFICA QUE:

El **Trabajo de Titulación Modalidad Proyectos de Investigación y Desarrollo**, denominado: Predicción del rendimiento académico por medio de técnicas de inteligencia artificial de los estudiantes de bachillerato en las Unidades Educativas fiscales del distrito educativo 05D06 (Salcedo - Ecuador), de responsabilidad del señor Klever Antonio Jijón Bautista, ha sido minuciosamente revisado por los miembros del Tribunal del trabajo de titulación, el mismo que cumple con los requisitos científicos, técnicos, legales, en tal virtud el Tribunal autoriza su presentación.

Dra. Martha Ximena Davalos Villegas, M.Sc.
PRESIDENTA



firmado electrónicamente por:
MARTHA XIMENA
DAVALOS VILLEGAS

Dr. Guillermo Edvin Machado Sotomayor, Ph. D.
DIRECTOR



firmado electrónicamente por:
GUILLERMO EDVIN
MACHADO SOTOMAYOR

Dr. Manuel Antonio Meneses Freire, Ph. D.
MIEMBRO



firmado electrónicamente por:
MANUEL ANTONIO
MENESES FREIRE

Lic. Nelly Narcisa Angulo Remache, Mgtr.
MIEMBRO



firmado electrónicamente por:
NELLY NARCISA
ANGULO REMACHE

Riobamba, julio 2023

DEDICATORIA

A mis padres Felix Antonio Jijón Mora y Miguelina de Lourdes Bautista Balarezo.

AGRADECIMIENTO

A Dios por los bienes materiales e inmateriales que me ha concedido sin haberlos merecido, por haberme dado los padres que me formaron y que siempre tuvieron en mente que el estudio nos permite crecer como personas.

A mis padres por dedicar su tiempo y trabajo para guiarme y formarme de la mejor manera, dejando de lado incluso el vestirse, el comer una golosina, con el objetivo que sus hijos podamos asistir a la Universidad.

A mi tutor y miembros del tribunal por su apoyo y trabajo incondicional para poder culminar este proyecto.

Antonio Jijón

TABLA DE CONTENIDO

RESUMEN.....	xiii
ABSTRACT.....	xiv
CAPÍTULO I.....	1
1 INTRODUCCIÓN.....	1
1.1 Formulación del problema	3
1.2 Justificación de la investigación.....	3
1.3 Objetivos	4
1.3.1 Objetivo general.....	4
1.3.2 Objetivos específicos	4
1.4 Hipótesis.....	4
CAPÍTULO II	5
2 MARCO TEÓRICO	5
2.1 Antecedentes del Problema.....	5
2.2 Bases Teóricas.....	6
2.2.1 Aprendizaje Automático	7
2.2.1.1 Técnicas de Aprendizaje Automático para regresión.....	8
2.2.2 Inteligencia Artificial (IA)	8
2.2.3 Regresión Lineal Múltiple.....	8
2.2.3.1 Finalidades de la (RLM)	9
2.2.3.2 Tipos de variable para regresión múltiple lineal (RLM).....	10
2.2.3.3 Número de sujetos y número de variables	10
2.2.3.4 Transformaciones para regresión múltiple lineal	11
2.2.3.5 Supuestos del modelo de regresión lineal múltiple	11
2.2.4 Regresión Logística.....	12
2.2.5 Rendimiento Académico	13

2.2.5.1	Características del Rendimiento Académico.....	15
2.2.5.2	Factores que inciden en el rendimiento académico.....	15
CAPÍTULO III.....		17
3	METODOLOGÍA DE INVESTIGACIÓN.....	17
3.1	Metodología	17
3.2	VARIABLES DE ESTUDIO	17
3.3	Adquisición de Datos	20
3.3.1	Población y muestra de estudio.....	20
3.4	Regresión Lineal Múltiple.....	22
3.4.1	Regresión Lineal Múltiple para instituciones fiscales rurales.....	23
3.4.2	Regresión Lineal Múltiple para instituciones fiscales urbanas	27
3.5	Regresión Logística Ordinal	30
3.5.1	Cambio de variable continua a variable nominal o factor.....	30
3.5.2	Regresión logística ordinal, instituciones fiscales rurales.....	31
3.5.3	Regresión logística ordinal, instituciones fiscales urbanas	37
CAPÍTULO IV.....		43
4	RESULTADOS Y DISCUSIÓN	43
4.1	Modelo de regresión lineal múltiple, instituciones fiscales rurales.....	43
4.1.1	Análisis de regresión lineal múltiple para instituciones educativas rurales	43
4.1.2	Predicción del rendimiento académico, instituciones educativas fiscales rurales.....	44
4.1.3	Modelo de regresión lineal múltiple para predecir el rendimiento académico de estudiantes de instituciones educativas fiscales rurales.	45
4.1.4	Validación del modelo de regresión lineal múltiple para instituciones educativas fiscales rurales.....	46
4.2	Modelo de regresión lineal múltiple, instituciones fiscales urbanas	47
4.2.1	Análisis de regresión lineal múltiple para instituciones educativas urbanas.....	47
4.2.2	Predicción del rendimiento académico, instituciones educativas fiscales urbanas	47
4.2.3	Modelo de regresión lineal múltiple para predecir el rendimiento académico de estudiantes de instituciones educativas fiscales urbanas.	48

4.2.4	Validación del modelo de regresión lineal múltiple para instituciones educativas fiscales urbanas.	49
4.2.5	Comparación entre Modelos de regresión lineal Múltiple obtenidos.....	50
4.3	Regresión logística ordinal, instituciones fiscales rurales.....	51
4.3.1	Análisis regresión logística ordinal, instituciones fiscales rurales	51
4.3.2	Matriz de confusión, instituciones rurales.....	52
4.4	Regresión logística ordinal, instituciones fiscales urbanas	52
4.4.1	Análisis regresión logística ordinal, instituciones fiscales urbanas	52
4.4.2	Matriz de confusión, instituciones urbanas	53
4.5	Planteamiento de la Hipótesis.	53
4.6	Prueba de Hipótesis.....	54
CAPÍTULO V		56
5	PROPUESTA.....	56
5.1	Justificación	56
5.2	Modelos matemáticos propuestos.....	56
5.2.1	Modelo de regresión lineal múltiple instituciones educativas fiscales urbanas.	56
5.2.2	Modelo de regresión lineal múltiple instituciones educativas fiscales rurales.....	57
5.3	Software a ser utilizado	57
5.4	Recursos necesarios	57
5.4.1	Recursos materiales y tecnológicos.....	57
5.4.2	Recursos humanos.....	58
5.4.3	Recurso tiempo	58
CONCLUSIONES		59
RECOMENDACIONES		60
GLOSARIO		
BIBLIOGRAFÍA		
ANEXOS		

ÍNDICE DE TABLAS

Tabla 1-3: Variables analizadas durante el estudio.....	17
Tabla 2-3: Población estudiantes instituciones fiscales rurales	20
Tabla 3-3: Población estudiantes instituciones fiscales urbanas.....	21
Tabla 4-3: Muestra estudiantes instituciones fiscales rurales	22
Tabla 5-3: Muestra estudiantes instituciones fiscales urbanas.....	22
Tabla 6-3: Resultados de regresión lineal múltiple, todas las variables, instituciones rurales...	23
Tabla 7-3: Regresión lineal múltiple solo con variables significativas, instituciones rurales....	25
Tabla 8-3: Variables seleccionadas con el criterio AIC, método paso a paso automático, instituciones rurales.....	26
Tabla 9-3: Resultados de regresión lineal múltiple, todas las variables, instituciones urbanas .	27
Tabla 10-3: Regresión lineal múltiple con variables significativas, instituciones urbanas.....	29
Tabla 11-3: Variables seleccionadas con el criterio AIC, método paso a paso automático, instituciones urbanas	29
Tabla 12-3: Cambio variable dependiente de continua a nominal.....	30
Tabla 13-3: Cambio variable dependiente de continua a nominal.....	31
Tabla 14-3: Primer resultado de regresión ordinal, instituciones rurales, casi todas las variables	31
Tabla 15-3: Variables significativas con p-valor de regresión ordinal rural, ingreso manual.....	33
Tabla 16-3: Predicción versus real, regresión ordinal criterio P-valor, instituciones rurales	34
Tabla 17-3: Matriz de confusión, regresión ordinal rural, criterio P-valor	34
Tabla 18-3: Variables seleccionadas con el criterio AIC, método paso a paso automático.....	35
Tabla 19-3: Predicción versus real, regresión ordinal con criterio AIC, instituciones rurales...	36
Tabla 20-3: Matriz de confusión, regresión ordinal rural, criterio AIC.....	36
Tabla 21-3: Primer resultado de regresión ordinal, instituciones urbanas, casi todas las variables	37
Tabla 22-3: Variables significativas con P-valor de regresión ordinal urbana, ingreso manual	38
Tabla 23-3: Predicción versus real, regresión ordinal con criterio P-valor, instituciones urbanas	39
Tabla 24-3: Matriz de confusión, regresión ordinal urbana, criterio P-valor.....	40
Tabla 25-3: Variables seleccionadas con el criterio AIC, método paso a paso automático.....	40
Tabla 26-3: Predicción versus real, regresión ordinal con criterio AIC, instituciones urbanas .	41
Tabla 27-3: Matriz de confusión, regresión ordinal urbana, criterio AIC.....	42
Tabla 1-4: Valores reales versus valores predichos, utilizando el modelo de regresión lineal múltiple en Rstudio, instituciones rurales	44
Tabla 2-4: Valores reales versus valores predichos, utilizando el modelo de regresión lineal múltiple en Rstudio, instituciones urbanas.....	48

ÍNDICE DE FIGURAS

Figura 1-4: Gráfica de residuos versus valores estimados de modelo lineal rural.....	45
Figura 2-4: Gráfica de residuos versus valores ajustados de modelo lineal urbano.	48

ÍNDICE DE ANEXOS

ANEXO A. Código de programación transformación a variables tipo factor.

ANEXO B. Código de programación regresión lineal múltiple: criterio AIC automático y criterio P-valor.

ANEXO C. Código de programación predicción utilizando modelo de regresión lineal múltiple seleccionado.

ANEXO D. Código de programación gráfico de predicción versus residuos.

ANEXO E. Código de programación Regresión Logística Ordinal criterio AIC automático y criterio P-valor.

ANEXO F. Código de programación Cambio de variables a variable tipo factor para predicción con regresión logística ordinal.

ANEXO G. Código de programación Predicción con regresión logística ordinal.

ANEXO H. Código de programación Matriz de confusión, porcentajes de error y acierto para regresión logística ordinal.

RESUMEN

El objetivo de esta investigación fue determinar los modelos matemáticos que predicen el rendimiento académico de estudiantes de bachillerato en instituciones educativas fiscales del cantón Salcedo. Las variables a estudiar se determinaron acudiendo a libros y artículos científicos donde indican los factores que influyen, predicen o explican el rendimiento académico, obteniendo 21 variables. Accedimos a información de 7 instituciones educativas fiscales rurales, 279 estudiantes y a 3 instituciones educativas fiscales urbanas, 402 estudiantes, mediante encuestas y las macros de notas; los datos fueron organizados y depurados mediante Excel. Se analizaron los datos mediante el software R, haciendo uso de la regresión lineal múltiple con el criterio del R-cuadrado ajustado y error estándar para determinar si los modelos son buenos predictores del rendimiento académico; por otra parte para la regresión logística se utilizó la matriz de confusión con porcentajes de aciertos para determinar si el modelo es un buen predictor. En la regresión lineal múltiple de instituciones rurales se determinaron 6 variables predictoras con un R-cuadrado ajustado de 0.8057 y un error estándar igual a 0.2373, en las instituciones urbanas se obtuvieron 7 variables, con un error estándar de 0.3 y un R cuadrado ajustado de 0.7525; en la regresión logística ordinal para las instituciones rurales se determinó un modelo con 5 variables predictoras y un porcentaje de acierto de 63.09524%, para las instituciones urbanas son 6 variables predictoras y un porcentaje de acierto de 52.45902%. Concluimos que la regresión logística no resultó una buena herramienta para predecir el rendimiento académico de los estudiantes de bachillerato de las instituciones educativas fiscales del cantón Salcedo, la regresión lineal múltiple si es una buena herramienta para predecir el rendimiento académico, lo han demostrado otras investigaciones y se corrobora lo mencionado con los resultados de esta investigación.

Palabras Clave: <REGRESIÓN LINEAL MÚLTIPLE>, <REGRESIÓN LOGÍSTICA>, <R-CUADRADO AJUSTADO>, < ERROR ESTÁNDAR >, <MATRIZ DE CONFUSIÓN>, <RENDIMIENTO ACADÉMICO>.



Desarrollado por:
LUIS ALBERTO
CAMINOS VARGAS



27-06-2023

0066-DBRA-UPT-IPEC-2023

ABSTRACT

The aim of this research was to determine the mathematic models that predict the academic performance of high school students in public educational institutions in Salcedo. The variables of the study were determined by books and academic papers which indicate the factors that affect, predict, or explain academic performance; obtaining 21 variables. We had access to the information of 7 public educational institutions in the countryside with a total of 279 students, and 3 public educational institutions in the city with a total of 402 students through surveys and their grades. Data were organized and elaborated by Excel. The information was analyzed by the software R; having used the multiple linear regression with the adjusting R square, and the standard error to determine which models are good indicators of academic performance. On the other hand, for the logistic regression, the matrix of confusion with the percentage of correct answers was used to determine if the model was a good indicator. In the multiple linear regression of the institutions in the countryside, it was determined 6 predicting variables with an adjusting R square of 0.8057, and a standard error of 0.2373. In the institutions of the city, it was obtained 7 variables with a standard error equal of 0.3, and an adjusting R square of 0.7525. In the logistic ordinal regression of the institutions in the countryside, it was determined a model with 5 predicting variables and a percentage of correct answers of 63.09524%. For the institutions of the city, there are 6 predicting variables and a percentage of 52.45902. We conclude that the logistic regression was not a good tool to predict the academic performance of high school students in public educational institutions in Salcedo. The multiple linear regression is a good tool for predicting academic performance, it is demonstrated through other research, and it is confirmed by the mentioned results of this investigation.

Key words: <MULTIPLE LINEAR REGRESSION>, <LOGISTIC REGRESSION>, <ADJUSTING R SQUARE>, < STANDARD ERROR >, <MATRIX OF CONFUSION>, <ACADEMIC PERFORMANCE>.

CAPÍTULO I

1 INTRODUCCIÓN

La investigación en el área de la educación despierta intereses en los actores que integran el sistema educativo del país, ya que son múltiples factores los involucrados en su estudio y la diferente forma en que estos factores influyen en el proceso educativo. En consecuencia, los resultados anteriores pueden ayudar a tomar medidas e implementar estrategias para favorecer la continuidad académica y reducir el fracaso y deserción estudiantil.

Las dificultades están presentes en todos los niveles educativos del sistema de educación, pues de alguna manera el rendimiento estudiantil y el desarrollo de conocimientos, social y cultural de los estudiantes son un aspecto importante.

El rendimiento estudiantil es una manera de medir el avance del estudiante, el resultado de sus esfuerzos en el marco de una mezcla de actividades y la apreciación más o menos acertada de los logros alcanzados (Baños, Ortiz, Baena, & Tristán, 2017).

Dificultades en el rendimiento escolar son una problemática de aprendizaje y pueden tener un origen no único estos están siempre presentes, en su mayoría son trabajables con el fin de mejorarlos, especialmente aquellos asociados a determinadas asignaturas o a contenidos concretos, estos también pueden ser asociados al nivel intelectual del estudiantado, o quizás con situaciones más complejas relacionadas con alteraciones neuropsicológicas, estas requieren de atención especializada, sin embargo, la mayoría tienen dificultades para resolver problemas, o no desarrollan estrategias de aprendizaje sumando a esto la falta de práctica, entonces al no intervenir podría generar un bajo rendimiento (Erazo, 2018).

En consecuencia, los problemas en el rendimiento escolar, generalmente son moderados y mejorables si logran ser abordados a tiempo, están caracterizados por expresar perturbaciones en el aprendizaje de ciertas áreas como las matemáticas, idiomas, y arrojar rendimientos por debajo de la capacidad real del estudiante, con cierta indisciplina escolar, desmotivación, debilidades en destrezas cognitivas y metacognitivas, entre muchos más, estas son evidentes en

la mayoría de fases escolares pero con prevalencia significativa en la adolescencia (Erazo, 2018).

Todas las instituciones educativas anterior y actualmente producen grandes cantidades de información que no son procesadas satisfactoriamente, es decir, no son utilizadas de la mejor manera; pues no se emplean herramientas tecnológicas que faciliten el adecuado procesamiento de la información. Las unidades educativas del Cantón Salcedo, generan información constantemente respecto a las actividades académicas y administrativas, en la que resalta la parte académica. Por ello, se ha planteado en la investigación la finalidad de implementación de técnicas de inteligencia artificial para los estudiantes de Bachillerato con el afán de predecir el rendimiento académico que tendrán los alumnos que cursan los últimos años de escolaridad, para dar paso a su primer año de la Universidad. El estudio facilitará el determinar o tener una estimación del éxito o fracaso que tendrán los estudiantes a nivel de bachillerato; información que permitirá a docentes y autoridades adoptar e implementar estrategias que mejoren el proceso de enseñanza de forma más personalizada, programas de ayuda y de seguimiento, apoyo psicológico e incentivos a aquellos estudiantes cuyo rendimiento estimado arroje resultados no satisfactorios, beneficiando no solo a los estudiantes.

Las técnicas de Inteligencia Artificial con las que se trabaja en este estudio son las técnicas basadas en regresión lineal múltiple y regresión logística que permita determinar un modelo matemático para clasificar a un nuevo estudiante en una de las categorías de rendimiento escolar. Esta clasificación permite a una institución educativa identificar con anticipación estudiantes con problemas potenciales de rendimiento académico, haciendo uso de software matemático para facilitar la obtención de la función que modela el rendimiento académico de los estudiantes de bachillerato de la institución.

Existen una gran variedad de estudios enfocados al rendimiento escolar, de los cuales se han determinado una gran cantidad de factores que intervienen en el mismo y una gran cantidad de clasificaciones, una de ellas indica que el rendimiento académico de un estudiante se ve influenciado por factores pedagógicos, físicos, sociodemográficos, psicosociales, institucionales y socioeconómicos. No obstante, el presente estudio se basará principalmente en los factores asociados al entorno en el que se desarrolla el estudiante, aquellos que refieren a los factores sociodemográficos y socioeconómicos enfocados al cantón Salcedo; por lo que como sector, interesa saber por qué los estudiantes rinden mejor que otros, considerando los factores mencionados.

En tal sentido se plantean interrogantes como la siguiente: ¿Es posible predecir el Rendimiento Académico, con la utilización de técnicas matemáticas y estadísticas de inteligencia artificial mediante funciones de regresión: lineales y logísticas, para los estudiantes de bachillerato de las unidades educativas fiscales del Distrito Educativo 05D06 (Salcedo-Ecuador), considerando sus factores sociodemográficos y socioeconómicos?

Para dar respuesta a esta interrogante se plantean los siguientes objetivos de investigación.

1.1 Formulación del problema

- ¿Es posible predecir el Rendimiento Académico, por medio de Técnicas de Inteligencia Artificial como la Regresión Lineal Múltiple y Regresión Logística de los estudiantes de bachillerato de las unidades educativas fiscales del Distrito Educativo 05D06 (Salcedo-Ecuador)?

1.2 Justificación de la investigación

La presente investigación toma importancia porque procura la calidad educativa ya que es uno de los objetivos fundamentales en los sistemas educativos y en los establecimientos académicos. Por consiguiente, se planifican e implementan estrategias con el fin de la mejora continua del rendimiento académico y la permanencia estudiantil en las instituciones educativas hasta culminar los ciclos educativos correspondientes.

Suma importancia ya que los resultados que se deriven de la investigación servirán como base para la toma de decisiones por parte de las Autoridades y personal Docente de las diferentes unidades educativas del cantón Salcedo, pues permitirá encontrar modelos matemáticos, predictivos del rendimiento escolar de los estudiantes de unidades educativas fiscales del cantón. Con los modelos encontrados se puede estimar el rendimiento académico estudiantil para un determinado estudiante y la probabilidad de éxito o fracaso durante el bachillerato general unificado (BGU), si en base al modelo encontrado se estima un bajo rendimiento académico de un estudiante, los docentes, autoridades, representantes y personal del departamento de consejería estudiantil pueden proponer actividades a fin de mejorar la participación del estudiante en el proceso educativo. Adicionalmente aporta nuevo conocimiento sobre los métodos de Modelamiento Matemático aplicados a las ciencias sociales, específicamente a la educación los cuales han sido poco estudiados formalmente a nivel de distritos educativos.

1.3 Objetivos

1.3.1 Objetivo general

Predecir el rendimiento académico por medio de técnicas de inteligencia artificial de los estudiantes de bachillerato de las unidades educativas fiscales del Distrito Educativo 05D06 (Salcedo-Ecuador) considerando sus factores sociodemográficos y socioeconómicos.

1.3.2 Objetivos específicos

- Realizar un diagnóstico y determinar los factores sociodemográficos y socioeconómicos que intervienen en el rendimiento académico mediante revisión de la literatura y opinión de expertos.
- Desarrollar el diseño del arte enfocado a la predicción del rendimiento académico mediante técnicas de inteligencia artificial para estudiantes de bachillerato en el cantón Salcedo.
- Proponer un modelo matemático que prediga el rendimiento escolar en bachillerato para las instituciones educativas fiscales urbanas y un modelo para las instituciones fiscales rurales del cantón salcedo, mediante la aplicación de Regresión Lineal Múltiple y Regresión Logística a partir de datos obtenidos de los estudiantes.
- Validar los modelos matemáticos propuestos, tanto para las instituciones educativas fiscales urbanas como para las instituciones educativas fiscales rurales del cantón salcedo y realizar un análisis comparativo entre modelos.

1.4 Hipótesis

La Inteligencia Artificial mediante Regresión Lineal Múltiple y Regresión Logística permiten predecir el Rendimiento Académico de los estudiantes de bachillerato de las unidades educativas fiscales del Distrito Educativo 05D06 (Salcedo-Ecuador).

CAPÍTULO II

2 MARCO TEÓRICO

2.1 Antecedentes del Problema

Castrillón, Sarache y Ruiz (2020), en su investigación cuyo tema es: Predicción del rendimiento académico con el uso de técnicas de inteligencia artificial, plantearon como objetivo fundamental estimar el rendimiento educativo de estudiantes de universidad partiendo de diferentes factores a través de las técnicas de inteligencia artificial. Una vez definidos los factores socioeconómicos influyentes: antecedentes familiares, educacionales, conductuales, etc., se diseña una metodología para clasificar a un estudiante de nuevo ingreso, en una de las cinco categorías e rendimiento académico. Dicha clasificación otorga a un determinado establecimiento universitario reconocer con anticipación a estudiantes con dificultades de aprendizaje. Gracias a este diagnóstico, pueden adoptarse intervenciones para una apropiada acción académica que permita mejorar la problemática encontrada. El proyecto se implementó en una institución universitaria pública colombiana, logrando como resultado un nivel de éxitos del 91.7%.

Rodríguez et al. (2018) En su estudio: causantes socio-demográficos que se relacionan al rendimiento educativo en la carrera de Médico Cirujano en Puebla, donde los factores estudiados fueron: sexo, edad, estado civil alumno, estado civil padres, número hermanos, ingreso familiar mensual, nivel educativo padres, número hijos (en caso de que el alumno tenga), si hace otra actividad a más de estudiar y promedio del nivel anterior. Se utilizó SPSS veintitresava versión donde aplica la estadística descriptiva para encontrar las medidas de tendencia central, dispersión. Para el análisis inferencial, se utiliza los estadísticos: T-Student en muestras independientes, para la asociación de rendimiento escolar con las variables cualitativas (dicotómicas). Para la asociación del rendimiento con las variables cualitativas con más de 2 factores el ANOVA. El coeficiente Pearson, para determinar la correlación con las variables cuantitativas. Finalmente para determinar la asociación entre variables cualitativas se utiliza el Chi-cuadrado. Se analizaron los datos de 192 estudiantes, 108 hombres y 84 mujeres con edades entre 18 y 29 años. En dicho estudio se obtuvieron los siguientes resultados: los factores socio-demográficos que permiten un mejor rendimiento académico, son: ser mujer, nivel académico alto madre y buen promedio del nivel previo. Trabajar, ingreso mensual familiar alto produce rendimiento escolar bajo.

Ríos (2010), en el desarrollo de su trabajo de investigación para obtener su título de Dra. En demografía realiza el trabajo denominado: Rendimiento Educativo versus causantes socio-demográficos. El estudio se realiza en la Ciudad de Córdoba para lo cual se tomó una cohorte estudiantil que ingresaron en el año 2000 con métodos de análisis multivariable, con la técnica de regresión lineal. El universo de estudio estaba compuesto por 2893 estudiantes de donde se obtuvo una muestra de 490 estudiantes para realizar el trabajo en el que se consideran las siguientes variables independientes: sexo, edad, estado conyugal, número de hijos, lugar de procedencia, situación de convivencia, situación laboral, nivel educativo de los padres, antecedentes de la escolaridad en secundaria. Como variables dependientes se considera el rendimiento académico basado en: avance de la carrera, promedio de todos los exámenes rendidos por el estudiante, índice de eficiencia que es la relación entre exámenes rendidos y exámenes aprobados. Los resultados obtenidos en el estudio arrojaron como variables significativas las siguientes: la situación de convivencia, localización de la institución secundaria, promedio de secundaria, nivel de estudios padre, nivel de estudios madre. Estas variables influyen en el rendimiento académico de los estudiantes.

Garbanzo (2014), realiza una investigación donde analiza las causas asociadas al rendimiento educativo considerando niveles socio-económicos, donde utiliza una regresión múltiple lineal con datos de estudiantes de la Universidad de Costa Rica. Algo muy importante que menciona en su estudio es que: hay múltiples factores que influyen en el rendimiento académico estas son de categoría sociodemográficas, psicosociales, pedagógicas e institucionales; es común encontrarse indicadores socioeconómicos dentro la categoría sociodemográfica. En este estudio se utilizan técnicas de recolección de datos y análisis cuantitativo. La variable dependiente es el promedio acumulado del estudiante hasta el momento de realizar la recolección de datos, se trabajó con la información de 120 estudiantes con una edad promedio de 23 años, dentro del trabajo con las variables categóricas se las transformó a variables dummy. Las variables que mostraron tener relación estadística significativa con el rendimiento académico fueron: acceso a beca, elección propia de carrera, conversación en casa sobre temas académicos, horas de trabajo, notas de admisión. Mientras que las variables resultan ser de importancia pero no son muy útiles para predecir el rendimiento académico son: sexo estudiante, edad, hábitos de estudio, lugar residencia, apoyo en proceso educativo, nivel estudios madre.

2.2 Bases Teóricas

El modelamiento matemático es una forma de enseñanza que permite relacionar el conocimiento matemático con la vida diaria, siendo las matemáticas la parte central en la

enseñanza y el aprendizaje. De esta manera, las actividades de modelización matemática facilitan el proceso de aprendizaje y permiten al estudiante aprender varios conceptos matemáticos que antes parecían complejos, además de algunas destrezas y herramientas para que sea posible el estudio y crítica de procesos que inclusive transforme a la modelización en un objetivo de la educación (Blum, 2003).

Luego de mencionar lo anterior, es muy importante comprender los diferentes conceptos que se mencionarán a lo largo de la investigación, como son el rendimiento educativo, la inteligencia artificial, la regresión múltiple y logística, entre otros subtemas de importantes para la investigación, mismos que se analizan a continuación.

2.2.1 Aprendizaje Automático

Para Miranda (2015) el Aprendizaje Automático es parte de la inteligencia computacional, que sirve para crear sistemas que pueden aprender autónomamente. Lo que se denomina aprendizaje consiste en la facilidad del sistema para encontrar una serie de patrones determinados por un gran número de parámetros. Una computadora no aprende por sí sola, sino un algoritmo de su programación, este se modifica constantemente con el ingreso de datos de esta manera puede predecir situaciones futuras o tomar acción automáticamente según las condiciones presentes. Debido a que estas acciones son realizadas de manera autónoma por el sistema, por esto se le conoce como aprendizaje automático, sin intervención de personas.

El algoritmo es un conjunto de códigos que incentivan al usuario a estudiar un conjunto complicado de datos hasta simplificarlo; cada algoritmo tiene varias instrucciones que llevan a lograr varios objetivos. A más de esto tienen la finalidad determinar y encontrar patrones, es decir, tendencias que faciliten la predicción. Los parámetros se obtienen de los datos conocidos como datos de entrenamiento; mientras más datos los resultados son más reales y se mejoran las predicciones. El aprendizaje puede ser supervisado, no supervisado o de refuerzo (Miranda, 2015).

Para Miranda (2015) el Análisis de Regresión es una subclasificación del Aprendizaje Autónomo computarizado conocido como aprendizaje supervisado automático. El propósito de este es establecer un patrón que relaciona un conjunto de características y una variable dependiente que puede ser continua o discreta.

2.2.1.1 Técnicas de Aprendizaje Automático para regresión

Para Miranda (2015) existen algunas técnicas automáticas de aprendizaje para estudios de regresión. Las más utilizadas son:

- a) Regresión lineal (variable con datos continuos).
- b) Regresión logística (variables con datos discretos).

2.2.2 Inteligencia Artificial (IA)

El desarrollo de la tecnología en los últimos tiempos genera gran influencia en la inteligencia computacional, esto va cambiando el razonamiento de las personas sobre la capacidad de hacer que las maquinas piensen por su cuenta; esto nos invita a reflexionar sobre la tecnología mundialmente. La inteligencia artificial es una parte de la informática encargada de diseñar un mundo centrado en programas y métodos que permitan visualizar la conducta inteligente de un objeto de estudio. La IA analiza cantidades enormes de datos, encontrar patrones o tendencias que permitan realizar predicciones de manera rápida y eficaz (Tironi & Cavallo, 2014).

La IA se encarga de realizar actividades específicas, donde los algoritmos usan técnicas innovadoras que promueven el cambio de la social de como la conocemos hacia un nuevo rumbo de estudios de ciencia ficción (programas informáticos) que solucionen problemas de razonamiento (Teigens, 2014).

Para Teigens (2014) algunos ejemplos de lo mencionado son:

- a. Siri. – Desempeñada como asistente personal la cual emplea un lenguaje entendible para todos los usuarios.
- b. Facebook. – actualmente reconoce la parte facial de una persona y da la alternativa de ser etiquetada.
- c. Amazon. – Da recomendaciones para canastas de compras.
- d. Waze. – Da información sobre el tráfico en una determinada ciudad.

2.2.3 Regresión Lineal Múltiple

La regresión lineal es una de las técnicas más utilizadas en el aprendizaje automatizado. En el modelo de regresión lineal, la variable dependiente o salida, debe ser continua. El modelo de regresión lineal se basa en el aprendizaje supervisado, lo cual significa que para conseguir el modelo matemático, debemos usar bases de datos con variables independientes, esto ayuda a predecir en el futuro, estos datos se los conoce como datos de entrenamiento (Granados, 2016).

Según Granados (2016) tenemos a disposición algunas técnicas de regresión y dependen del tipo de variable y la relación entre ellas. Las más utilizadas son las regresiones lineales, estas suponen que existe una relación lineal de la variable dependiente con las variables independientes. Las regresiones lineales pueden ser simples, que relaciona dos variable una dependiente y una independiente, también tenemos las regresiones lineales múltiples, en el modelo de regresión lineal múltiple, tenemos una sola variable dependiente y varias independientes, por ejemplo el peso de alguna persona depende de la edad, altura, género y otros. Para Granados (2016) el modelo de regresión lineal múltiple arroja como resultado una función del tipo:

$$y_j = b_0 + b_1x_{1j} + b_2x_{2j} + \dots + b_kx_{kj} + u_j, \text{ para } j=1, 2, 3 \dots n$$

En forma matricial:
$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{k1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \dots & x_{kn} \end{bmatrix} \begin{bmatrix} b_0 \\ \vdots \\ b_k \end{bmatrix} + \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$$

$$y = xb + u$$

Donde **y** se la denomina variable **endógena**, variable dependiente, entre otros; **x** son variables a los cuales se las llama: **exógena**, independiente; **u** conocidos como residuos y **b** coeficientes estimados (Granados, 2016).

2.2.3.1 Finalidades de la (RLM)

Si trabajamos con Regresión lineal múltiple un objetivo es encontrar una función matemática. Según Morales (2011), el uso más generalizado de las ecuaciones de regresión múltiple lineal es la determinación de las variables predictoras más eficaces, esto da una guía más práctica que de análisis solamente teórico. También hay que considerar que no solamente se busca determinar las mejores predicciones, a más de eso se busca analizar el grupo de variables que tienen lógica y sentido al formar la ecuación, estas deben ser elegidas con argumento fundamentado.

Uso para explicación:

Cuando buscamos un modelo o ecuación para la explicación, decimos que trabajamos en la parte teórica, en el entendimiento de las causantes que generan un determinado comportamiento de las variables, no en predecir dicha variable, además se realiza la interpretación de los coeficientes beta de la función matemática, hay que ver la correlación de cada variable independiente con la variable dependiente, estas son más de información (Morales, 2011).

Uso para predicción:

Al trabajar con la regresión múltiple lineal para predicción, es importante también tomar en cuenta la ecuación obtenida de la regresión con los coeficientes beta para cada variable independiente en este tipo de uso de la regresión lineal múltiple es de suma importancia quitar de la ecuación las variables que aportan poco o nada a la predicción (Morales, 2011).

2.2.3.2 Tipos de variable para regresión múltiple lineal (RLM)

Continua: Las variables continuas son aquellas que ocupan todo el espacio de números. Son números reales, es decir enteros, racionales con decimales o sin decimales. generalmente son variables de cantidad como el peso o la edad (Granados, 2016).

Discretas: Las discretas se representan con números enteros. Además, las variables suelen ser **cualitativas**, utilizadas para identificar o codificar una característica (género, tipo auto, etc.) si la alternativa de respuesta tiene solo 2 opciones, entonces se las conocen como **dicotómicas** (tiene internet). Si tienen varias alternativas de respuesta, más de 2, entonces se las conoce como **factor**. A las variables exógenas de tipo factor y de tipo ordenadas se las puede transformar en **dummies** de ceros y unos (Granados, 2016).

2.2.3.3 Número de sujetos y número de variables

Hay varios autores cuya recomendación es obtener de 10 a 20 veces el número de sujetos sometidos a estudio que variables independientes, si no se cumple esto las correlaciones múltiples no son las adecuadas. Cuando las hipótesis están muy claras, datos precisos, y no hay valores perdidos, entonces a veces es suficiente con cinco objetos de estudio por variable, 20 objetos de estudio por cada variable independiente es adecuado en cualquier circunstancia. Además de lo mencionado anteriormente es fundamental tener muy en cuenta los elementos atípicos que se puedan presentar ya que estos pueden afectar de manera negativa los coeficientes

de regresión que se pretenden obtener; por lo tanto es recomendable dejar de lado aquellos sujetos que presenten datos extremos o atípicos (Morales, 2011).

2.2.3.4 Transformaciones para regresión múltiple lineal

En la Regresión Múltiple Lineal su característica fundamental es que la relación entre variables dependiente e independiente debe ser lineal; al momento de realizar el estudio o investigación con varios datos no siempre se obtiene una relación lineal entre dichas variables de estudio, en ese caso se debe manejar la linealización o transformación para lograr asociaciones entre variables y que estas asociaciones o relaciones sean lineales. Si el trabajo de análisis se realiza con variables continuas o discretas las técnicas más utilizadas son el trabajo con logaritmos o trabajar con polinomios (Granados, 2016).

2.2.3.5 Supuestos del modelo de regresión lineal múltiple

Según Baños, Fonseca, Álvarez (2019) el proceso natural para obtener el modelo de regresión múltiple lineal involucra algunas suposiciones que en posterior se deben comprobar, estas suposiciones se muestran a continuación:

- a) Para Baños, Fonseca, Álvarez (2019) Linealidad es la relación de variable dependiente y las variables independientes debe ser lineal.

$$E(u_j) = 0$$

- b) Para Baños, Fonseca, Álvarez (2019) La Independencia consiste en que los errores de las variables independientes no sean dependientes entre ellos; entonces, cada u_i es independiente para $(j = 1, 2, 3, \dots, n)$
- c) Para Baños, Fonseca, Álvarez (2019) La Homocedasticidad consiste en que la varianza de los errores debe ser constante, es decir:

$$Var(u_j) = 0$$

- d) Para Baños, Fonseca, Álvarez (2019) en La Normalidad, las variables planteadas en el estudio deben estar apegadas o seguir la ley normal, es decir:

$$u_j \varepsilon N(0, \sigma^2)$$

- e) Para Baños, Fonseca, Álvarez (2019) la No colinealidad consiste en la no existencia de correlación entre variables independientes.

2.2.4 *Regresión Logística*

Para Sánchez (2000) en la regresión logística a diferencia de la regresión lineal, se predice una variable nominal y no una continua, es decir, en la regresión logística se predice categorías (bueno, malo, regular). Las regresiones logísticas se utilizan en múltiples estudios tales como el diagnóstico médico, para determinar si un paciente puede o no adquirir una enfermedad; o en el marketing para saber el color que prefieren las mujeres al comprar un determinado producto (rosa, rojo, blanco).

La regresión múltiple lineal ha tenido varios usos hasta el momento, sin embargo, el poder explicativo de la técnica es escaso, salvo algunas excepciones, entonces esto ha llevado a trabajar o investigar con nuevas técnicas, una de estas alternativas nos permite pronosticar si un estudiante tendrá éxito o si este fracasará en el proceso estudiantil (García et al., 2000).

La regresión logística, una técnica muy utilizada para predecir del rendimiento escolar de estudiantes (Reyes, 2003). Esta técnica facilita la determinación de la probabilidad de que ocurra un evento, basados en variables conocidas como predictoras. La regresión lineal y la regresión logística mide y valora cuan influyente es cada variable independiente con respecto a la variable dependiente, además controla el efecto que tienen entre variables independientes. (Sánchez, 2000).

Villarruel et al. (2020) estudió la influencia de los factores: fisiológicos, socio-económicos, culturales, académicos, demográficos, y psicológicos, en el rendimiento escolar de los estudiantes, donde utilizan la regresión logística, este muestra el grado de relación de las variables significativas que determinan el rendimiento académico de los estudiantes de secundaria ecuatorianos.

Para García et al. (2000) si analizamos una variable que da como resultado una respuesta dicotómica (si, no o ganar, perder) y queremos analizar los efectos que algunas variables tienen sobre esta (independientes), entonces la regresión binaria logística es la técnica recomendable pues se puede:

- a) determinar estimadamente la probabilidad que se dé un evento esperado (tener éxito en primer año de bachillerato), conociendo los valores que toman las variables independientes en la toma de datos,
- b) Evaluar la influencia que cada variable independiente tiene sobre la respuesta en forma de (OR) u (ODD) RATIO. Si la (OR) RATIO es mayor que uno, esto indica mayor probabilidad que ocurra el evento, y si la (OR) RATIO es menor que uno, entonces indica una menor probabilidad de ocurrencia.

Según García et al. (2000) la Regresión Logística, nos ayuda a estudiar la relación funcional entre una variable dependiente con varias categorías (nominal) (Y) y un grupo de “n” variables independientes $X = (X_{1i}, X_{2i}, \dots, X_{kj})$ estas variables independientes pueden ser cuantitativas continuas, cuantitativas discretas o categóricas.

Según García et al. (2000) el modelo o ecuación de regresión logística se escribe de la siguiente manera:

$$\log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{kj}$$

Dónde p es la probabilidad de ocurrencia del evento esperado (en nuestro caso tener éxito en el bachillerato) a partir del conjunto de variables independientes que se plantean en el estudio.

Para García et al. (2000) el modelo de una regresión logística dicotómica o binaria permite predecir utilizando la probabilidad de éxito del evento de interés.

De esta manera se obtiene la ecuación:

$$P(Y = 1/X) = \frac{e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{kj}}}{1 + e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{kj}}}$$

El proceso de la estimación o inferencia con regresión logística, se consigue al trabajar con el vector de datos X_0^m en la ecuación estimada; para predecir el rendimiento escolar de un estudiante (Éxito o fracaso), Según García et al. (2000) la ecuación será de la siguiente manera:

$$P(Y = 1/X_0^m) = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_k X_{kj}}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_k X_{kj}}}$$

En este caso se hace también el uso de la significancia de los coeficientes de cada variable del modelo ayudándonos del estadístico de Wald Lehmann, similarmente al modelo de regresión múltiple lineal. Es decir, si una variable independiente luego del análisis no es significativa, entonces deberíamos quitarle y no formaría parte del modelo (García et al., 2000).

2.2.5 Rendimiento Académico

Para Reyes (2018) la educación es un hecho deliberado, si nos referimos a la calidad en la educación, la gran mayoría de sistemas educativos buscan mejorar el rendimiento escolar de los estudiantes, dando paso a una evolución constante año tras año. Entonces, cuando se da un

cambio, esto permite iniciar un análisis de cómo se lleva la educación y las técnicas que se aplican para llevarlas de la mejor manera, en este sentido la variable dependiente generalmente utilizada es el rendimiento escolar, este se define de la siguiente manera:

Rendimiento viene del latín *reddere* que significa: recuperación o pago. La relación entre logros alcanzados, lo mencionado representa el nivel de éxito en la universidad, en el deporte, el trabajo, etc.; utilizando como ejemplo la educación, entonces nos permite entender que la relación del trabajo que realizan los docentes y el trabajo que realizan los estudiantes nos permitirá entender mediante trabajo científico el desempeño escolar, por un lado nos ayuda a entender la educación, el trabajo que se realiza en la investigación sobre el desempeño por otro lado, y en conjunto se deben tomar en cuenta los factores que influyen en el desempeño (Reyes, 2018).

El desempeño escolar en un entorno educativo, donde los profesores dan mucha importancia al trabajo duro. En otras palabras, el estudiante espera ser reconocido por su habilidad (aspecto importante de la autoestima), y sus esfuerzos en el aula. En conclusión, el desempeño de la escuela está siempre en movimiento, no es estático, siempre en respuesta al proceso de enseñanza-aprendizaje, entonces, está relacionado directamente con la capacidad del estudiante y el esfuerzo o trabajo que este le dedique al proceso educativo (Goleman, 2008).

Mejorar el rendimiento escolar es desarrollar las habilidades del estudiantado y mejorar las habilidades psicológicas de los mismos, mediante el uso de diversos métodos en el proceso de enseñanza-aprendizaje. Para valorar el rendimiento escolar o logro académico se suele medirlo durante un período de tiempo, este generalmente se lo realiza mediante evaluaciones al final del periodo mencionado.

El rendimiento escolar podría ser analizado y entendido como el valor mínimo obtenido en la escala de aprobación y el nivel máximo obtenido en la escala de desaprobación ante ciertos conocimientos, es decir mide la capacidad de respuesta del estudiante y es personal, individual; es decir, se mide lo que tiene un estudiante de forma estimada y lo que ha aprendido durante el proceso educativo, esta es una meta o propósito previamente establecido (Martínez & Pérez, 1996).

El rendimiento escolar puede estar relacionado según el modelo académico; este modelo puede juzgar o medir el nivel de logro de los objetivos de formación planteados, también se puede evaluar los factores que directamente se relacionen con el desempeño escolar, estos factores pueden ser: la familia y su relación con el estudiante, la sociedad y como el estudiante la percibe, las actividades que el estudiante realiza extracurricularmente y el mismo entorno estudiantil en el que se desempeña (Medina, 2011).

En conclusión: el rendimiento escolar se presenta como la forma de medir la capacidad de responder o la capacidad de ser instruido un estudiante, arrojando como resultado de manera estimada lo que un estudiante ha aprendido en el proceso enseñanza-aprendizaje, estos dependen de las metas u objetivos establecidos en un sistema educativo.

2.2.5.1 Características del Rendimiento Académico

De acuerdo a Martínez & Pérez (1996) que basados en las distintas formas de definir desempeño escolar, concluye que existe una clasificación estática y otra dinámica que involucra al estudiante como alguien social en la educación. Principalmente este autor considera que el rendimiento escolar tiene las siguientes características:

- a) Desde una visión dinámica, el rendimiento escolar está relacionado con el proceso de aprendizaje, es decir, con la capacidad del alumno y el esfuerzo del mismo.
- b) Desde una visión estática, incluye resultados producidos por el estudiante y conductas de logro que este presenta.
- c) El desempeño escolar también se relaciona con la medición de los estándares de la calidad y el juicio que se tiene de la evaluación.
- d) La actuación es un medio, y no es un fin;
- e) El desempeño escolar se relaciona con propósitos y metas éticas, esto incluye lo económico, entonces mide un desempeño escolar basado en modelos de sociedad actuales.

2.2.5.2 Factores que inciden en el rendimiento académico

Según Garbanzo (2014) el rendimiento universitario estudiantil es importante para evaluar el estándar educativo, es el resultado de la interacción de diferentes elementos que se relacionan en el desempeño académico de la población estudiantil. Lo anteriormente mencionado es visible por medio de las calificaciones obtenidas por cada estudiante en el proceso, generalmente los resultados se obtiene de forma cuantitativa. Las calificaciones resultado de las evaluaciones representan el nivel en el desempeño durante la escolaridad, el proceso educativo involucra una gran cantidad de variables, estas pueden ser personales, académicas, sociales, estas están interrelacionadas entre unas con otras. También mencionan que las variables intervinientes en el rendimiento son: las pedagógicas, institucionales, sociodemográficas y psicosociales.

Los factores asociados a las instituciones educativas se refieren a características estructurales y funcionales, estas son horarios de trabajo por curso, involucran también aspectos relacionados con la carrera elegida; finalmente también considera el ambiente en la institución. Dentro los factores pedagógicos encontramos el trabajo docente, las relaciones docente-alumno, destrezas del docente para comunicarse con los estudiantes, la habilidad para despertar el interés del estudiante por la educación, la capacidad para acceder a los estudiantes y el acceso que este brinde para que los estudiantes accedan al docente en términos de educación y confianza, para esto el docente debe valorarse y valorar al estudiante persona en su totalidad. Los recursos didácticos que utilice y ponga a disposición del estudiantado, técnicas y métodos que el docente utilice para evaluar aprendizajes, etc. (Garbanzo, 2014).

Entre los factores socio-demográficos se toman en cuenta aquellos relacionados a la economía de la familia del estudiante, su entorno socio-económico y socio-cultural, sexo, tipo de institución según su clasificación a la que el estudiante acceda, escolaridad de los representantes del estudiante. Los factores asociados a la parte cultural y social de la familia se consideran importante en el desarrollo del intelecto del estudiante. Finalmente los factores socio-económicos se relacionan con los bienes o pertenencias a disposición del grupo familiar (Garbanzo, 2014).

Es visible que el rendimiento académico es multicausal, intervienen en él factores de diversa índole, estos deben ser considerados: factores psicosociales, sociodemográficos, pedagógicos, socioeconómicos e institucionales.

CAPÍTULO III

3 METODOLOGÍA DE INVESTIGACIÓN

3.1 Metodología

Este capítulo está dedicado a indicar el proceso y pasos que se siguen para determinar los modelos matemáticos que nos permiten predecir el rendimiento académico de los estudiantes de bachillerato de las instituciones educativas fiscales rurales y urbanas. Se inicia el trabajo tomando las variables que han sido estudiadas en otros proyectos de investigación realizados y que han resultado significativas para predecir el rendimiento académico, estos estudios se presentan en el capítulo 2, los antecedentes de la investigación, luego se procede a la recolección de datos de los estudiantes de las instituciones educativas del cantón Salcedo. Con el software R-studio se hace el análisis de los datos mediante regresión lineal múltiple y regresión logística ordinal para determinar los modelos que predicen el rendimiento académico y verificar si la predicción es considerada buena.

3.2 Variables de estudio

Como se mencionó anteriormente las variables seleccionadas para el estudio se basan en investigaciones realizadas anteriormente, factores que inciden en el rendimiento académico en otros países, en Ecuador, a nivel de bachillerato y superior; es importante mencionar que estas variables tienen un enfoque en los factores socio-económicos y socio-demográficos. Las variables seleccionadas se muestran en la siguiente tabla.

Tabla 1-3: Variables analizadas durante el estudio

Código	Descripción	Tipo Variable	Niveles Variable
Y2	Promedio promoción primero BGU.	Continua	de 7,00 a 10,00
Y3	Promedio promoción primero BGU.	Cualitativa (ordinal)	0: promedios de 7,00 - 7,50 1: promedios mayores de 7,50 a 8,00 2: promedios mayores de 8,00 a 8,50 3: promedios mayores de 8,50 a 9,00

			4: promedios mayores de 9,00 a 9,50 5: promedios mayores de 9,50 a 10,00 de 7,00 a 10,00
X1	Promedio básica superior (8vo, 9no, 10mo)	Continua	
X2	Promedio promoción octavo EGB.	Continua	de 7,00 a 10,00
X3	Promedio promoción noveno EGB.	Continua	de 7,00 a 10,00
X4	Promedio promoción décimo EGB.	Continua	de 7,00 a 10,00
X5	Institución Educativa	Cualitativa (nominal)	1: U.E. LUIS ALFREDO MARTÍNEZ 2: U.E. EMILIO TERÁN 3: U.E. CUSUBAMBA 4: U.E. PAPAHERCO 5: U.E. JOSÉ EMILIO ÁLVAREZ 6: U.E. NÉSTOR MOGOLLÓN 7: U.E.I.B. ALICIA MARCUATH 8: U.E. ROSA ZÁRATE 9: U.E. SALCEDO 10: U.E. 19 DE SEPTIEMBRE
X6	Tipo Institución	Cualitativa (dicotómica)	1: urbana 2: rural
X7	Género	Cualitativa (dicotómica)	1: femenino 2: masculino
X8	Edad	Cualitativa (nominal)	1: 16 años 2: 17 años 3: 18 años 4: 19 años 5: 20 años o más
X9	Escolaridad representante (hombre)	Cualitativa (nominal)	1: ninguna 2: básico 3: bachillerato 4: técnico, policía, militar 5: universitario
X10	Escolaridad representante (mujer)	Cualitativa (nominal)	1: ninguna 2: básico 3: bachillerato 4: técnico, policía, militar 5: universitario

X11	Título que aspiran del estudiante	Cualitativa (nominal)	1: bachiller 2: policía, militar 3: técnico 4: universitario
X12	Interés representantes proceso educativo	Cualitativa (nominal)	1: nada 2: muy poco 3: poco 4: mucho
X13	Estado civil estudiante	Cualitativa (dicotómica)	1: casado, unión libre 2: soltero
X14	Responsables del hogar	Cualitativa (nominal)	1: papá y mamá 2: padrastro y mamá 3: madrastra y papá 4: solo papá 5: solo mamá 6: otros familiares, otros
X15	Ingreso mensual al hogar	Cualitativa (nominal)	1: 200 o menos 2: mayor a 200, menor a 400 3: mayor a 400, menor a 600 4: mayor a 600, menor 1000 5: 1000 o más
X16	Número de personas en el hogar	Cualitativa (nominal)	1: una persona 2: 2 personas 3: 3,4 o 5 personas 4: 6,7,8 o 9 personas 5: 10 o más personas
X17	Relación familiar	Cualitativa (nominal)	1: mala 2: regular 3: buena 4: muy buena
X18	Tiempo para llegar a la institución	Cualitativa (nominal)	1: 5 o 10 minutos 2: 20 o 30 minutos 3: 40, 50 o 60 minutos 4: entre 75 y 90 minutos 5: 2 horas o más.
X19	Tipo de transporte	Cualitativa (nominal)	1: a pie o bicicleta 2: mayor tiempo a pie y auto 3: mayor tiempo en auto y a pie 4: medio de transporte
X20	Servicio de internet	Cualitativa (dicotómica)	1: no 2: si

Realizado por: Jijón, A. 2023

En la tabla se muestran los tipos de variables utilizadas como dependientes, independientes, cualitativas, cuantitativas, continuas, discretas, ordinales, nominales. La variable Y2, es de tipo continua, es el promedio de promoción de primero de bachillerato, ella es nuestra variable dependiente para la regresión lineal múltiple, la variable Y3 es una variable nominal ordenada con varios factores que se obtiene a partir de la variable Y2. Las variables X1, X2, X3... son nuestras variables independientes tanto para la regresión lineal múltiple como para la regresión logística ordinal de las instituciones rurales y urbanas.

3.3 Adquisición de Datos

Para adquirir o recoger los datos se utilizan encuestas para todas las variables distintas a calificaciones o promedios de promoción, mientras que para las variables que involucran las calificaciones de los estudiantes se recurrió a solicitar al Distrito Educativo Salcedo 05D06 las macros, que son archivos digitalizados en formato Excel donde constan los promedios de promoción de los estudiantes desde octavo de educación general básica hasta tercero de bachillerato general unificado.

3.3.1 Población y muestra de estudio

La población considerada para el presente estudio corresponde a un total de 836 estudiantes que en el momento de recolectar los datos cursan el tercer año de bachillerato en las unidades educativas fiscales del Distrito Educativo 05D06 (Salcedo-Ecuador). La clasificación de acuerdo a la unidad educativa es como sigue:

Tabla 2-3: Población estudiantes instituciones fiscales rurales

Institución Educativa	Número estudiantes
U.E. CUSUBAMBA	54
U.E.I.B. ALICIA MARCUATH DE YEROVI	9
U.E. JOSÉ EMILIO ÁLVAREZ	21
U.E. PPAHURCO	25
U.E. LUIS ALFREDO MARTÍNEZ	118
U.E. NÉSTOR MOGOLLÓN	16
U.E. EMILIO TERÁN	55
Total	298

Realizado por: Jijón, A. 2023

Tabla 3-3: Población estudiantes instituciones fiscales urbanas

Institución Educativa	Número estudiantes
U.E. 19 DE SEPTIEMBRE	214
U.E. ROSA ZÁRATE	119
U.E. SALCEDO	205
Total	538

Realizado por: Jijón, A. 2023

Se buscó realizar el levantamiento de datos con toda la población, dado que es una población limitada relativamente corta y la información debe obtenerse de las respectivas instituciones, sin embargo, por aspectos asociados a la pandemia existieron datos que no se pudieron alcanzar, por lo que se analiza a continuación el cálculo de la muestra para determinar el mínimo de estudiantes que se deberá disponer para el levantamiento de datos, de acuerdo a cada grupo. Se utiliza la fórmula del tamaño de muestras finitas de población:

$$n = \frac{N z_{\alpha/2}^2 P(1 - P)}{(N - 1)e^2 + z_{\alpha/2}^2 P(1 - P)}$$

Dónde:

N: Corresponde al tamaño de población, en cada caso.

Z_{α/2}: Corresponde al valor Z, es decir la cantidad de desviaciones estándar en relación a la media. Se utiliza en este caso un Z de 1,96 correspondiente a un 95% de confianza para el estudio.

P: Representa la proporción de aceptación de la población. Para este estudio la aplicabilidad requiere un valor de P de 0,5 con el cual los resultados obtenidos tendrán mayor confiabilidad.

e: Representa el error permisible, el cual se ha seleccionado en un +/-5% que se puede asumir dentro del estudio.

Por tanto el valor de cada muestra para el grupo de instituciones educativas rurales será:

$$n = \frac{298(1,96)^2 0,5(1 - 0,5)}{(298 - 1)0,05^2 + 1,96^2(0,5)(1 - 0,5)}$$

$$n = 168,07 \approx \mathbf{169}$$

Para el grupo de instituciones educativas urbanas es:

$$n = \frac{538(1,96)^2 0,5(1 - 0,5)}{(538 - 1)0,05^2 + 1,96^2(0,5)(1 - 0,5)}$$

$$n = 224,37 \approx \mathbf{225}$$

Tabla 4-3: Muestra estudiantes instituciones fiscales rurales

INSTITUCIÓN EDUCATIVA	N° de		Cantidad de muestra
	estudiantes	Porcentaje	
U.E. CUSUBAMBA	54	18,12080537%	31
U.E.I.B. MARCUATH DE YEROVI	9	3,020134228%	5
U.E. JOSÉ EMILIO ÁLVAREZ	21	7,046979866%	12
U.E. PAPAURCO	25	8,389261745%	14
U.E. LUIS ALFREDO MARTÍNEZ	118	39,59731544%	67
U.E. NÉSTOR MOGOLLÓN	16	5,369127517%	9
U.E. EMILIO TERÁN	55	18,45637584%	31
Total 1	298		169

Realizado por: Jijón, A. 2023

Tabla 5-3: Muestra estudiantes instituciones fiscales urbanas

INSTITUCIÓN EDUCATIVA	N° de		Cantidad de muestra
	estudiantes	Porcentaje	
U.E. 19 DE SEPTIEMBRE	213	39,77695167%	89
U.E. ROSA ZÁRATE	120	22,11895911%	50
U.E. SALCEDO	187	38,10408922%	86
Total 2	538		225
TOTAL GENERAL (1+2)	836		384

Realizado por: Jijón, A. 2023

Ahora es importante mencionar que se logró recolectar la información de un total de 402 estudiantes de las instituciones fiscales urbanas, por encima de la cantidad necesaria; este patrón se repite para cada institución educativa. En el caso de las instituciones educativas fiscales rurales se logró recolectar la información de 278 estudiantes, nuevamente por encima de la muestra requerida.

3.4 Regresión Lineal Múltiple

3.4.1 Regresión Lineal Múltiple para instituciones fiscales rurales

Como se menciona anteriormente el software utilizado es R-studio, en el cual se introducen las variables seleccionadas para el estudio mediante una tabla de datos. La función utilizada para la regresión lineal múltiple es (lm), cabe recalcar que se trabaja con la introducción de variables en la regresión y basados en aquellas que resultan significativas se consideran para el próximo nivel en donde se utilizan solo aquellas variables significativas; también se utiliza la introducción de variables paso a paso automático hacia adelante y hacia atrás, el cual selecciona el conjunto de variables que menor valor (AIC) arroje, y la función utilizada es (stepAIC).

Tabla 6-3: Resultados de regresión lineal múltiple, todas las variables, instituciones rurales

Variables	Coefficientes	Error estándar	T valor	P valor	Significancia
Intercepto					
Intercepto	1,4846383	0,4759797	3,119	0,00205	**
X2	0,0418887	0,0672287	0,623	0,53386	
X3	0,2474476	0,0749004	3,304	0,00111	**
X4	0,5490967	0,0547014	10,038	< 2e-16	***
X52	-0,2161606	0,0474521	-4,555	8,57E-06	***
X53	-0,065852	0,0582809	-1,13	0,25972	
X54	0,0684376	0,0655392	1,044	0,2975	
X55	-0,0726377	0,0684562	-1,061	0,28979	
X56	-0,2482148	0,0827333	-3	0,003	**
X57	-0,22677	0,1021763	-2,219	0,02746	*
X72	-0,0966344	0,0331077	-2,919	0,00387	**
X82	0,0793711	0,0936419	0,848	0,39756	
X83	0,1120256	0,0947386	1,182	0,23827	
X84	0,0165327	0,112648	0,147	0,88345	
X85	0,0038146	0,1737979	0,022	0,98251	
X92	-0,0185472	0,0563443	-0,329	0,74233	
X93	-0,0331991	0,0630381	-0,527	0,59895	
X94	-0,2366142	0,2708862	-0,873	0,38333	
X102	-0,0245491	0,0565596	-0,434	0,66468	
X103	-0,0550581	0,0653769	-0,842	0,40059	
X104	0,2059899	0,1994814	1,033	0,30289	
X105	0,1165867	0,1736568	0,671	0,50268	
X112	-0,0033278	0,0469489	-0,071	0,94356	
X113	-0,0430849	0,0689867	-0,625	0,53291	
X114	0,1024556	0,041725	2,455	0,01483	*
X122	-0,1194111	0,2787121	-0,428	0,66874	
X123	0,1168139	0,2605789	0,448	0,65438	
X124	0,0625914	0,2566658	0,244	0,80756	
X132	-0,1791896	0,1129472	-1,586	0,11403	

X142	-0,093742	0,084199	-1,113	0,26675
X143	0,404012	0,2531231	1,596	0,11187
X144	-0,0443891	0,0435374	-1,02	0,30903
X145	-0,1602087	0,1414384	-1,133	0,25854
X146	-0,1672602	0,1269047	-1,318	0,18884
X152	-0,0017789	0,0385052	-0,046	0,96319
X153	-0,046403	0,0435696	-1,065	0,288
X154	0,1833125	0,108467	1,69	0,09241 .
X155	0,0811857	0,0976213	0,832	0,40649
X162	-0,1860886	0,1911947	-0,973	0,33145
X163	-0,2438877	0,1576909	-1,547	0,12336
X164	-0,2795615	0,1583822	-1,765	0,0789 .
X165	-0,263247	0,1707601	-1,542	0,12457
X172	0,0018701	0,1317972	0,014	0,98869
X173	0,0310482	0,1250642	0,248	0,80416
X174	0,0285502	0,1256756	0,227	0,82049
X182	0,0009412	0,039023	0,024	0,98078
X183	-0,0637878	0,0536577	-1,189	0,23578
X184	0,1408608	0,1174336	1,199	0,2316
X185	0,105151	0,1903082	0,553	0,58113
X192	-0,0115487	0,0569107	-0,203	0,83938
X193	-0,0035156	0,0561326	-0,063	0,95012
X194	-0,0380054	0,0514668	-0,738	0,46101
X202	0,022015	0,0418034	0,527	0,59897

Error estándar de residuos: 0,2395; R² ajustado: 0,802

Realizado por: Jijón, A. 2023

En la tabla se puede observar que R-studio tiene la ventaja que por sí solo transforma los factores o variables nominales en variables dummy de ceros y unos, por esta razón el programa arroja resultados para las variables y cada variable analizada para cada factor que la compone; por ejemplo: X9 (Escolaridad del representante) con sus alternativas correspondientes, 1: ninguno; 2: básico; 3: bachillerato; 4: policía, militar, técnico; 5: universitario. Se puede ver los siguientes códigos o variables: **X92** indica que la representante ha estudiado solo hasta el subnivel básico de escolaridad; **X93** indica que la representante ha estudiado hasta el nivel de bachillerato y así sucesivamente; además se puede ver que no se considera el código **X91** pues la opción 1: ninguna escolaridad de la representante toma el programa como referencia. Además en este punto es importante indicar que las variables X3, X4, X5, X7, X11 son significativas, también lo es el intercepto, para determinar si son o no significativas nos fijamos en que el P-valor sea menor a 0,05; además el programa utiliza una marca para decirnos si es o no significativa la variable, todas aquellas que tienen 1, 2 o 3 asteriscos (*, **, ***) indican que las variables son de utilidad para predecir el rendimiento académico, mientras que las variables que tienen como marca un punto (.), no son significativas pero están cercanas. Consideramos un

valor de $\alpha = 0,05$ para determinar si una variable es significativa, es decir, si el P valor es menor que α , entonces se rechaza la hipótesis nula ($H_0: \beta_j = 0$, el coeficiente es cero) y se acepta la hipótesis alternativa ($H_1: \beta_j \neq 0$, el coeficiente es diferente de cero); Ejemplo: el P valor del intercepto es 0,00205 que es menor a $\alpha = 0,05$, entonces se rechaza la hipótesis nula a favor de la hipótesis alternativa y concluimos diciendo que el intercepto es significativo, aporta a la predicción.

Tabla 7-3: Regresión lineal múltiple solo con variables significativas, instituciones rurales

Variables	Coefficientes	Error estándar	T valor	P valor	Significancia
(Intercept)	1,1909195	0,2466778	4,828	2,35E-06	***
X3	0,2726728	0,0536697	5,081	7,18E-07	***
X4	0,5742237	0,0484383	11,855	< 2e-16	***
X52	-0,2058176	0,0418235	-4,921	1,53E-06	***
X53	-0,0639365	0,047607	-1,343	0,180436	.
X54	0,1088936	0,0584126	1,864	0,063414	.
X55	-0,0588109	0,0641047	-0,917	0,35977	.
X56	-0,1873119	0,0685378	-2,733	0,006706	**
X57	-0,1811521	0,0954182	-1,899	0,058732	.
X72	-0,0801192	0,030378	-2,637	0,008856	**
X112	0,018687	0,0424851	0,44	0,66041	.
X113	-0,0028814	0,064395	-0,045	0,964345	.
X114	0,1308653	0,0374422	3,495	0,000557	***
X152	0,0215741	0,0354097	0,609	0,542875	.
X153	-0,0001069	0,0383486	-0,003	0,997778	.
X154	0,2306691	0,1021805	2,257	0,024806	*
X155	0,0525584	0,0849851	0,618	0,536823	.

Residual standard error: 0,2373; Adjusted R-squared: 0,8056

F-statistic: 72,75 on 16 and 261 DF, p-value: < 2,2e-16

Realizado por: Jijón, A. 2023

En la tabla anterior se puede observar que las variables X3, X4, X5, X7, X11, resultaban significativas, sin embargo, para el siguiente paso se considera también aquellas variables que casi eran significativas y al momento de considerarlas con aquellas significativas logramos que también la variable X15 resultara significativa y diera al modelo mejores valores de R-cuadrada y reducir el error estándar. En la información que muestra en la tabla se considera solo las variables que resultan significativas en alguno de sus factores y estas son X3: promedio promoción noveno, X4: promedio promoción décimo, X5: institución educativa a la que pertenece, X7: género, X11: título que aspiran del estudiante, X15: ingreso mensual. Se puede también observar que el error estándar de la regresión lineal múltiple es de 0,2373; además nos

da un R-cuadrado ajustado de 0,8056, esto significa que las variables mencionadas explican un 81% del rendimiento académico Y2, promedio promoción de primero de bachillerato. Finalmente trabajando con la prueba F se puede observar que su valor es: 72,75 y tiene un P valor de 2,2e-16 que es menor que 0,05 de alfa, por lo tanto rechazamos la hipótesis nula: ninguna variable predice el rendimiento académico y aceptamos la hipótesis alternativa: al menos una variable predice el rendimiento académico.

Tabla 8-3: Variables seleccionadas con el criterio AIC, método paso a paso automático, instituciones rurales

Variables	Coefficientes
Intercepto	
Intercepto	1,27521
X3	0,27166
X4	0,56714
X52	-0,20942
X53	-0,06113
X54	0,09364
X55	-0,06896
X56	-0,19888
X57	-0,19559
X72	-0,08548
X112	0,02885
X113	-0,01039
X114	0,13539
X132	-0,18506
Residual standard error: 0,2373;	
Adjusted R-squared: 0,8057	
AIC= -786.25	
Realizado por: Jijón, A. 2023	

En la tabla se puede observar que utilizando el método de introducción de variables paso a paso en ambas direcciones, es decir, hacia adelante y hacia atrás, el programa nos indica que las variables: X3, X4, X5, X7, X11, X13, tienen la capacidad de predecir el rendimiento académico (Y2) de los estudiantes de las instituciones educativas fiscales rurales del cantón Salcedo, seleccionando el conjunto de variables que arrojan el menor valor del criterio AIC que en este caso es: -786.25. Ahora es importante mencionar que la variable X15 no es significativa en ninguno de sus factores. Si comparamos los valores del Error estándar y R^2 de los 2 criterios para las instituciones fiscales rurales, vemos que no hay mayor diferencia.

3.4.2 Regresión Lineal Múltiple para instituciones fiscales urbanas

De manera similar al modelo obtenido para las instituciones educativas rurales, en este caso se trabaja con la introducción de variables en la regresión y basados en aquellas que resultan significativas se consideran para el próximo nivel en donde se utilizan solo aquellas variables significativas; también se utiliza la introducción de variables paso a paso automático hacia adelante y hacia atrás, el cual selecciona el conjunto de variables que menor valor (AIC) arroje, y la función utilizada es (stepAIC).

Tabla 9-3: Resultados de regresión lineal múltiple, todas las variables, instituciones urbanas

Variab	Coefficientes	Error estándar	T valor	P valor	Significancia
Intercepto	1,556506	0,409944	3,797	0,000172	***
X2	-0,049924	0,047823	-1,044	0,297228	
X3	0,122108	0,065759	1,857	0,064156	.
X4	0,728578	0,061406	11,865	< 2e-16	***
X59	-0,091632	0,045414	-2,018	0,044376	*
X510	-0,111655	0,043691	-2,556	0,011019	*
X72	-0,049934	0,033699	-1,482	0,139292	
X82	0,020612	0,132007	0,156	0,876007	
X83	-0,011542	0,133638	-0,086	0,93122	
X84	-0,080154	0,145623	-0,55	0,582379	
X85	-0,14067	0,169817	-0,828	0,408023	
X92	0,019015	0,068387	0,278	0,781133	
X93	0,031995	0,072118	0,444	0,657567	
X94	-0,052397	0,160677	-0,326	0,744539	
X95	-0,103632	0,103616	-1	0,317922	
X102	-0,092185	0,06086	-1,515	0,130737	
X103	-0,048891	0,066222	-0,738	0,460826	
X104	0,007898	0,099981	0,079	0,937081	
X105	0,112548	0,083485	1,348	0,178483	
X112	0,017783	0,062947	0,283	0,777713	
X113	-0,023501	0,09508	-0,247	0,804917	
X114	0,024398	0,059886	0,407	0,683956	
X122	0,196252	0,257666	0,762	0,446775	
X123	0,164329	0,234187	0,702	0,483327	
X124	0,209863	0,231765	0,905	0,365817	
X132	0,232529	0,11418	2,037	0,042443	*
X142	0,087427	0,091064	0,96	0,337684	
X144	0,011914	0,046309	0,257	0,797123	
X145	0,115573	0,103584	1,116	0,265289	
X146	-0,055554	0,116627	-0,476	0,634126	
X152	0,013536	0,046867	0,289	0,772888	
X153	-0,017505	0,047738	-0,367	0,714068	

X154	-0,061711	0,068425	-0,902	0,367733
X155	0,077612	0,076627	1,013	0,311824
X163	-0,035924	0,1078	-0,333	0,739143
X164	-0,026102	0,111812	-0,233	0,815554
X165	-0,107717	0,156395	-0,689	0,49143
X172	0,011464	0,157031	0,073	0,941846
X173	0,003901	0,150256	0,026	0,9793
X174	-0,005801	0,151035	-0,038	0,969382
X182	-0,022754	0,037542	-0,606	0,544848
X183	0,055741	0,057482	0,97	0,332851
X184	0,167909	0,144731	1,16	0,246771
X192	-0,028038	0,054539	-0,514	0,607509
X193	-0,011551	0,05655	-0,204	0,838267
X194	-0,020921	0,047261	-0,443	0,658283
X202	0,082282	0,037251	2,209	0,027826 *

Error estándar de residuos: 0,3051; R² ajustado: 0,7444

Realizado por: Jijón, A. 2023

En la tabla se puede observar que Rstudio tiene la ventaja que por sí solo transforma los factores o variables nominales en variables dummy de ceros y unos, por esta razón el programa arroja resultados para las variables y cada variable analizada para cada factor que la compone; por ejemplo: X9 (Escolaridad del representante) con sus alternativas correspondientes, 1: ninguno; 2: básico; 3: bachillerato; 4: policía, militar, técnico; 5: universitario. Se puede ver los siguientes códigos o variables: **X92** indica que la representante ha estudiado solo hasta el subnivel básico de escolaridad; **X93** indica que la representante ha estudiado hasta el nivel de bachillerato y así sucesivamente; además se puede ver que no se considera el código **X91** pues la opción 1: ninguna escolaridad de la representante toma el programa como referencia. Para las instituciones urbanas las variables X3, X4, X5, X13, X20 y el intercepto son significativas, para determinar si son o no significativas nos fijamos en que el P valor sea menor a 0,05; además el programa utiliza una marca para decirnos si es o no significativa la variable, todas aquellas que tienen 1, 2 o 3 asteriscos (*, **, ***) indican que las variables son de utilidad para predecir el rendimiento académico, mientras que las variables que tienen como marca un punto (.), no son significativas pero están cercanas. Consideramos un valor de $\alpha = 0,05$ para determinar si una variable es significativa, es decir, si el P valor es menor que α , entonces se rechaza la hipótesis nula ($H_0: \beta_j = 0$, el coeficiente es cero) y se acepta la hipótesis alternativa ($H_1: \beta_j \neq 0$, el coeficiente es diferente de cero); Ejemplo: el P valor del intercepto es 0,00205 que es menor a $\alpha = 0,05$, entonces se rechaza la hipótesis nula a favor de la hipótesis alternativa y concluimos diciendo que el intercepto es significativo, aporta a la predicción.

Tabla 10-3: Regresión lineal múltiple con variables significativas, instituciones urbanas

Variables	Coefficientes	Error estándar	T valor	P valor	Significancia
(Intercept)	1,43389	0,21293	6,734	5,83E-11	***
X3	0,11324	0,05575	2,031	0,04288	*
X4	0,71533	0,05542	12,908	< 2e-16	***
X59	-0,10544	0,04175	-2,526	0,01193	*
X510	-0,13377	0,03903	-3,428	0,00067	***
X132	0,2223	0,10875	2,044	0,0416	*
X202	0,10145	0,03168	3,202	0,00147	**

Residual standard error: 0,303; Adjusted R-squared: 0,7475

F-statistic: 198,8 on 16 and 261 DF, p-value: < 2,2e-16

Realizado por: Jijón, A. 2023

En la información que muestra en la tabla se considera solo las variables que resultan significativas en alguno de sus factores y estas son X3: promedio promoción noveno, X4: promedio promoción décimo, X5: institución educativa a la que pertenece, X13: estado civil del estudiante, X20: si el estudiante dispone o no de servicio de internet. Se puede también observar que el error estándar de la regresión lineal múltiple es de 0,303; además nos da un R cuadrado ajustado de 0,7475 que es un valor aceptable para predecir el rendimiento académico de los estudiantes, esto significa que las variables mencionadas explican un 75% del rendimiento académico Y2, promedio promoción de primero de bachillerato. Finalmente trabajando con la prueba F se puede observar que su valor es: 198,8 y tiene un P valor de 2,2e-16 que es menor que 0,05 de alfa, por lo tanto rechazamos la hipótesis nula: ninguna variable predice el rendimiento académico y aceptamos la hipótesis alternativa: al menos una variable predice el rendimiento académico.

Tabla 11-3: Variables seleccionadas con el criterio AIC, método paso a paso automático, instituciones urbanas

Variables	Coefficientes
Intercepto	
intercepto	1,62257
X3	0,10625
X4	0,70709
X59	-0,10434
X510	-0,11493
X72	-0,04926
X102	-0,07328
X103	-0,02734
X104	0,05037
X105	0,10788

X132	0,22628
X202	0,08009
Residual standard error: 0,3;	
Adjusted R-squared: 0,7525	
AIC=-956,18	

Realizado por: Jijón, A. 2023

En la tabla se puede observar que utilizando el método de introducción de variables paso a paso en ambas direcciones, es decir, hacia adelante y hacia atrás, el programa nos indica que las variables: X3, X4, X5, X7, X10, X13, X20, tienen la capacidad de predecir el rendimiento académico (Y3) de los estudiantes de las instituciones educativas fiscales urbanas del cantón Salcedo, seleccionando el conjunto de variables que arrojan el menor valor del criterio AIC que en este caso es: -956,18. Ahora es importante mencionar que las variables X7 y X10 no son significativas en ninguno de sus factores pero comparando los valores de errores estándar se reducen de 0,303 con el criterio del P-valor a 0,3 con el criterio AIC y el R² ajustado en cambio aumenta de 0,7475 a 0,7525. Lo que indica que con el criterio AIC se obtiene un mejor conjunto de variables para predecir el rendimiento académico.

3.5 Regresión Logística Ordinal

3.5.1 Cambio de variable continua a variable nominal o factor

Para poder trabajar la regresión logística, es necesario que nuestra variable dependiente que es de tipo continua, promedio de promoción de primero de bachillerato tome una condición de variable nominal para lo cual vemos conveniente trabajarla de la siguiente manera:

Tabla 12-3: Cambio variable dependiente de continua a nominal

Intervalo Y2	Y3
7,00 <= Y2 <= 7,50	0
7,50 < Y2 <= 8,00	1
8,00 < Y2 <= 8,50	2
8,50 < Y2 <= 9,00	3
9,00 < Y2 <= 9,50	4
9,50 < Y2 <= 10,00	5

Realizado por: Jijón, A. 2023

En la tabla se puede observar que el cambio de la variable dependiente que es el promedio de promoción de primero de bachillerato se lo hace a variable nominal que toma valores en este caso numéricos, pero en la programación se la introduce como factor ordenado, ordenado debido a que lleva un orden ya que un promedio de 9,00 es mayor que un promedio de 7,54; es decir, importa la ponderación que estos tienen al momento de evaluarlos. El mismo cambio se hace tanto para los datos de los estudiantes de las instituciones educativas fiscales rurales y urbanas. Por ejemplo si un estudiante de la Unidad Educativa Luis A. Martínez obtuvo un promedio de promoción en primero de bachillerato Y2 igual a 8,75; nosotros le asignamos un valor Y3 de 3 que en la programación este 3 es considerado como texto o cualidad.

3.5.2 Regresión logística ordinal, instituciones fiscales rurales

Para obtener el modelo como se mencionó anteriormente es importante que nuestra variable continua sea transformada en factor o variable nominal y a esta de la misma forma introducirla al programa como variable ordinal, es decir, ordenada.

Tabla 13-3: Cambio variable dependiente de continua a nominal

Intervalo Y2	Y3
7,00 <= Y2 <= 7,50	0
7,50 < Y2 <= 8,00	1
8,00 < Y2 <= 8,50	2
8,50 < Y2 <= 9,00	3
9,00 < Y2 <= 9,50	4
9,50 < Y2 <= 10,00	5

Realizado por: Jijón, A. 2023

Luego de este cambio se introducen, analizan las variables y se obtienen los siguientes resultados.

Tabla 14-3: Primer resultado de regresión ordinal, instituciones rurales, casi todas las variables

Variable	Coefficiente	Error Estándar	Valor t	Valor p	Significancia
X2	0,2650947	5,86E-01	4,52E-01	6,51E-01	
X3	2,1129238	6,77E-01	3,12E+00	1,81E-03	**
X4	4,8162456	5,89E-01	8,18E+00	2,83E-16	***
X52	-1,67690578	4,42E-01	-3,79E+00	1,50E-04	***
X53	-0,27613347	5,43E-01	-5,09E-01	6,11E-01	

X54	0,97418449	5,88E-01	1,66E+00	9,73E-02	
X55	-0,83273861	6,28E-01	-1,33E+00	1,85E-01	
X56	-1,05207918	7,60E-01	-1,38E+00	1,66E-01	
X57	-1,8603946	9,09E-01	-2,05E+00	4,06E-02	*
X82	0,48826318	8,49E-01	5,75E-01	5,65E-01	
X83	0,92793645	8,62E-01	1,08E+00	2,82E-01	
X84	-0,73036273	1,05E+00	-6,95E-01	4,87E-01	
X85	-0,24998529	1,57E+00	-1,59E-01	8,73E-01	
X92	-0,10292761	5,00E-01	-2,06E-01	8,37E-01	
X93	-0,20550816	5,69E-01	-3,61E-01	7,18E-01	
X94	-12,3978325	9,59E-05	-1,29E+05	0,00E+00	***
X102	-0,16104718	5,08E-01	-3,17E-01	7,51E-01	
X103	-0,27759601	5,80E-01	-4,79E-01	6,32E-01	
X104	-0,81091188	1,64E+00	-4,93E-01	6,22E-01	
X105	-1,12616422	1,42E+00	-7,93E-01	4,28E-01	
X112	-0,31409518	4,21E-01	-7,46E-01	4,56E-01	
X113	-0,4940343	6,43E-01	-7,69E-01	4,42E-01	
X114	0,42630941	3,82E-01	1,12E+00	2,64E-01	
X122	-3,25999499	2,68E+00	-1,22E+00	2,24E-01	
X123	0,10215758	2,52E+00	4,06E-02	9,68E-01	
X124	-0,45518875	2,49E+00	-1,83E-01	8,55E-01	
X132	-0,20843833	1,09E+00	-1,91E-01	8,49E-01	
X152	0,02400083	3,46E-01	6,93E-02	9,45E-01	
X153	0,03325209	3,74E-01	8,89E-02	9,29E-01	
X154	1,71689275	1,04E+00	1,65E+00	9,83E-02	
X155	1,28277671	8,09E-01	1,59E+00	1,13E-01	
X162	-1,51929039	1,61E+00	-9,42E-01	3,46E-01	
X163	-1,19997996	1,23E+00	-9,78E-01	3,28E-01	
X164	-1,28102009	1,24E+00	-1,04E+00	3,00E-01	
X165	-1,3319391	1,35E+00	-9,88E-01	3,23E-01	
X182	-0,05058286	3,51E-01	-1,44E-01	8,86E-01	
X183	-0,7272688	4,91E-01	-1,48E+00	1,38E-01	
X184	0,67698611	1,02E+00	6,65E-01	5,06E-01	
X185	0,47050942	1,47E+00	3,19E-01	7,49E-01	
X192	-0,26574959	4,85E-01	-5,47E-01	5,84E-01	
X193	-0,00860621	5,00E-01	-1,72E-02	9,86E-01	
X194	-0,41281658	4,56E-01	-9,05E-01	3,65E-01	
X202	0,16056904	3,67E-01	4,38E-01	6,62E-01	

Realizado por: Jijón, A. 2023

En la tabla se puede observar los resultados que se obtuvieron en primera instancia para determinar el modelo utilizando regresión logística ordinal, basada en los p-valores. Se puede observar que en la tabla no se considera las variables X7, X14, X17, pues al considerarlas con todas las variables el programa nos arrojaba un mensaje de advertencia que decía, que se están obteniendo probabilidades de 0 o 1, por esta razón en primera instancia no se las considera.

Luego se hace el proceso de ir ingresando las variables ordenadamente y eliminando o no considerando aquellas que no resultan significativas, incluyendo las variables mencionadas. Basados en la tabla se puede decir que hasta el momento se consideran significativas las variables: X3, X4, X5, X9.

Tabla 15-3: Variables significativas con p-valor de regresión ordinal rural, ingreso manual

Variable	Coefficiente	Error Estándar	Valor t	Valor p	Significancia
X3	1,88904134	5,15E-01	3,67E+00	2,46E-04	***
X4	4,58635111	5,49E-01	8,36E+00	6,23E-17	***
X52	-1,4281273	3,90E-01	-3,67E+00	2,47E-04	**
X53	-0,52483433	4,36E-01	-1,20E+00	2,29E-01	
X54	1,02103997	5,11E-01	2,00E+00	4,55E-02	*
X55	-0,50281272	5,90E-01	-8,53E-01	3,94E-01	
X56	-0,96235143	6,12E-01	-1,57E+00	1,16E-01	
X57	-1,53114164	8,46E-01	-1,81E+00	7,04E-02	
X72	-0,5774285	2,72E-01	-2,12E+00	3,37E-02	*
X92	0,02715817	4,42E-01	6,14E-02	9,51E-01	
X93	-0,06737842	4,77E-01	-1,41E-01	8,88E-01	
X94	-10,4860739	5,59E-05	-1,88E+05	0,00E+00	***
X112	0,05361049	3,68E-01	1,46E-01	8,84E-01	
X113	-0,38215592	6,04E-01	-6,33E-01	5,27E-01	
X114	0,69854831	3,38E-01	2,07E+00	3,88E-02	*
0 1	48,3602721	3,99E+00	1,21E+01	7,78E-34	***
1 2	51,8408679	4,19E+00	1,24E+01	3,68E-35	***
2 3	55,9557248	4,46E+00	1,25E+01	4,21E-36	***
3 4	58,5268503	4,61E+00	1,27E+01	5,89E-37	***
4 5	62,4386495	4,82E+00	1,30E+01	1,97E-38	***

Realizado por: Jijón, A. 2023

Para obtener los resultados que se muestran en la tabla, como ya se mencionó anteriormente se realizó el ingreso de las variables manualmente de una en una, si la variable es significativa para el modelo se la sigue manteniendo, caso contrario ya no se la toma en cuenta, en otras palabras se está utilizando el método paso a paso, pero se lo realiza manualmente. Entonces se obtiene que las variables: X3, X4, son variables continuas significativas y las variables: X5, X7, X9, X11, son nominales y también son significativas en alguno de sus factores y serán consideradas para el modelo. Finalmente también los interceptos, valores de α (0|1, 1|2...) resultan todos significativos.

Tabla 16-3: Predicción versus real, regresión ordinal criterio P-valor, instituciones rurales

N°	Y3 Predicho	Y3 Real
1	0	0
2	1	0
3	1	1
4	1	0
5	1	1
6	2	2
7	1	1
8	2	1
9	1	0
10	0	0
11	2	2
12	2	2
13	2	2
14	2	2
15	2	2
16	3	3
17	4	4

Realizado por: Jijón, A. 2023

En la tabla se muestra un resumen de 17 predicciones realizadas con el modelo obtenido con el criterio P-valor para los datos de 17 estudiantes de las instituciones educativas fiscales rurales, se puede ver que mayormente coinciden los valores predichos con los valores reales.

Tabla 17-3: Matriz de confusión, regresión ordinal rural, criterio P-valor

Predicción	0	1	2	3	4
0	10	7	0	0	0
1	5	16	6	0	0
2	0	7	20	6	0
3	0	0	0	5	0
4	0	0	0	0	2
5	0	0	0	0	0

Realizado por: Jijón, A. 2023

En la tabla se puede observar la denominada matriz de confusión de los datos, es decir con esta tabla podemos darnos cuenta de cuantas respuestas son correctas y cuantas están erradas. Por ejemplo, si analizamos cuando Y3 vale 0 el modelo predice diez veces correctamente, esto se indica en la primera celda pintada, mientras que siete veces predice incorrectamente, es decir, en

vez de predecir 0, predice siete veces 1; de igual manera cuando Y3 es 1, se puede ver que el modelo predice correctamente dieciséis veces correctamente, la segunda celda pintada, mientras que predice once veces incorrectamente, cinco veces predice 0 y seis veces predice 2. Teniendo en consideración los errores y aciertos se llega a la conclusión que con este modelo se predice correctamente el 63,09524% y se predice incorrectamente el 36,90476.

Tabla 18-3: Variables seleccionadas con el criterio AIC, método paso a paso automático

Variable	Coficiente
X3	1,809299
X4	4,9207571
X52	-1,6544783
X53	-0,7820858
X54	0,8135913
X55	-0,8347683
X56	-1,0059369
X57	-1,8040209
X72	-0,6808085
X122	-2,1828971
X123	0,6467512
X124	-0,123741
0 1	49,89444
1 2	53,34002
2 3	57,50415
3 4	60,10054
4 5	64,1194

AIC: 473,567

Realizado por: Jijón, A. 2023

Para determinar las variables que influyen en el rendimiento académico clasificado como factores y estos factores son ordenados, en este caso se utiliza el método basado en la clasificación AIC, este método selecciona el conjunto de variables que mejor se ajustan, predicen la variable Y3 automáticamente, comparando los valores AIC para cada conjunto de variables analizadas, en este caso el programa selecciona el conjunto de variables que arrojan el menor valor de AIC (criterio de información de Akaike): 477,7481; dentro de este proceso se seleccionó el método paso a paso de introducción de variables hacia adelante y hacia atrás (both). Como resultado se obtiene que las Variables independientes: X3, X4, X5, X7, X12 están relacionadas con la variable dependiente Y3. Si de estas variables obtenemos su P valor resultan casi todas significativas excepto X12.

Tabla 19-3: Predicción versus real, regresión ordinal con criterio AIC, instituciones rurales

N°	Y3 Predicho	Y3 Real
1	0	0
2	1	0
3	1	1
4	1	0
5	1	1
6	2	2
7	1	1
8	2	1
9	0	0
10	0	0
11	2	2
12	2	2
13	2	2
14	2	2
15	2	2
16	3	3
17	2	3

Realizado por: Jijón, A. 2023

En la tabla se muestra un resumen de 17 predicciones realizadas con el modelo obtenido con el criterio AIC con los datos de 17 estudiantes de las instituciones educativas fiscales rurales, se puede ver que mayormente coinciden los valores predichos con los valores reales. Además es importante mencionar que para la predicción con el criterio del p-valor y el criterio AIC se utilizan los mismos valores reales que se muestran en las tablas correspondientes.

Tabla 20-3: Matriz de confusión, regresión ordinal rural, criterio AIC

Predicción	0	1	2	3	4
0	11	6	1	0	0
1	4	17	7	0	0
2	0	7	18	6	0
3	0	0	0	5	0
4	0	0	0	0	2
5	0	0	0	0	0

Realizado por: Jijón, A. 2023

En la tabla se puede observar la denominada matriz de confusión de los datos, es decir con esta tabla podemos darnos cuenta de cuantas respuestas son correctas y cuantas están erradas. Por

ejemplo, si analizamos cuando Y3 vale 0 el modelo predice once veces correctamente, esto se indica en la primera celda pintada, mientras que siete veces predice incorrectamente, es decir, en vez de predecir 0, predice seis veces 1 y una vez 2; de igual manera cuando Y3 es 1, se puede ver que el modelo predice diecisiete veces correctamente bien, la segunda celda pintada, mientras que predice once veces incorrectamente, cuatro veces predice 0 y siete veces predice 2. Teniendo en consideración los errores y aciertos se llega a la conclusión que con este modelo se predice correctamente el 63,09524% y se predice incorrectamente el 36,90476%. Que es igual a los porcentajes de acierto y error del modelo obtenido con el criterio p-valor.

3.5.3 Regresión logística ordinal, instituciones fiscales urbanas

De manera similar que en el caso de las instituciones educativas fiscales rurales, en este caso se realiza una transformación de la variable continua Y2, promedio de promoción de primero de bachillerato a variable nominal ordenada, variable ordinal; luego se introduce la matriz de datos al programa y se obtienen los siguientes resultados:

Tabla 21-3: Primer resultado de regresión ordinal, instituciones urbanas, casi todas las variables

Variable	Coefficiente	Error Estándar	Valor t	Valor p	Significancia
X3	0,83776724	0,3941697	2,12539717	3,36E-02	*
X4	4,29952902	0,4557537	9,43388702	3,95E-21	***
X59	-0,7549838	0,29974	-2,5187953	1,18E-02	*
X510	-0,86558548	0,2932575	-2,95162249	3,16E-03	**
X72	-0,21799205	0,2225158	-0,97967014	3,27E-01	
X92	0,03508491	0,4412284	0,07951642	9,37E-01	
X93	0,24719023	0,4682342	0,52792007	5,98E-01	
X94	-0,55516367	1,0854335	-0,51146725	6,09E-01	
X95	-0,69436107	0,6690619	-1,03781288	2,99E-01	
X102	-0,51273744	0,4010906	-1,27835827	2,01E-01	
X103	-0,00654072	0,4356998	-0,01501199	9,88E-01	
X104	0,5632124	0,6955564	0,80972926	4,18E-01	
X105	0,78208996	0,5432124	1,43974995	1,50E-01	
X112	0,32208968	0,4158337	0,77456363	4,39E-01	
X113	0,06822499	0,6162597	0,11070818	9,12E-01	
X114	0,18362825	0,3923811	0,46798448	6,40E-01	
X122	1,57006388	1,8374499	0,85447985	3,93E-01	
X123	1,88249422	1,7032819	1,10521588	2,69E-01	
X124	1,91385206	1,6861485	1,13504363	2,56E-01	
X132	1,01882641	0,7187372	1,41752283	1,56E-01	
X142	0,32618433	0,5894987	0,55332494	5,80E-01	

X144	0,36622958	0,309525	1,18319881	2,37E-01
X145	0,86756351	0,6751794	1,28493774	1,99E-01
X146	-0,43525076	0,73429	-0,59275049	5,53E-01
X152	0,16064946	0,3083927	0,5209249	6,02E-01
X153	0,02012484	0,3164462	0,0635964	9,49E-01
X154	-0,25384582	0,4416675	-0,57474415	5,65E-01
X155	0,9156359	0,504824	1,81377238	6,97E-02
X163	0,22298948	0,7106473	0,31378362	7,54E-01
X164	0,46700393	0,7383173	0,63252471	5,27E-01
X165	-0,20099895	1,0080061	-0,19940251	8,42E-01
X172	0,01548595	0,9949945	0,01556386	9,88E-01
X173	-0,15622012	0,9495414	-0,16452164	8,69E-01
X174	-0,15074756	0,9546763	-0,15790437	8,75E-01
X182	-0,0709967	0,2481482	-0,28610607	7,75E-01
X183	0,51779354	0,3748317	1,38140278	1,67E-01
X184	1,74966182	0,9913496	1,76492908	7,76E-02
X192	-0,18002896	0,3612649	-0,49832955	6,18E-01
X193	0,1493206	0,372913	0,40041668	6,89E-01
X194	0,06217813	0,3149837	0,19740109	8,44E-01
X202	0,50544754	0,2457934	2,05639196	3,97E-02 *

Realizado por: Jijón, A. 2023

En la tabla se puede observar los resultados que se obtuvieron en primera instancia para determinar el modelo utilizando regresión logística ordinal, basada en los P-valores. Se puede observar que en la tabla no se considera la variable X8, pues al considerarla con todas las variables el programa nos arrojaba un mensaje de advertencia que decía, que se están obteniendo probabilidades de 0 o 1, por esta razón en primera instancia no se la considera. Luego se hace el proceso de ir ingresando las variables ordenadamente y eliminando o no considerando aquellas que no resultan significativas, incluyendo la variable mencionada. Basados en la tabla se puede decir que hasta el momento se consideran significativas las variables: X3, X4, X5, X20.

Tabla 22-3: Variables significativas con P-valor de regresión ordinal urbana, ingreso manual

Variable	Coefficiente	Error Estándar	Valor t	Valor p	Significancia
X3	0,7931219	0,3661812	2,165928	3,03E-02	*
X4	4,0518616	0,4236854	9,563373	1,14E-21	***
X59	-0,7497764	0,2763103	-2,713531	6,66E-03	**
X510	-0,9195576	0,2619183	-3,510856	4,47E-04	***
X202	0,5304392	0,2095442	2,531395	1,14E-02	*
0 1	35,3706888	2,3023804	15,36266	2,91E-53	***

1 2	38,4610537	2,4139425	15,932879	3,75E-57	***
2 3	41,162322	2,5474205	16,158432	9,90E-59	***
3 4	44,2013679	2,6982621	16,381421	2,60E-60	***
4 5	46,8202767	2,7965264	16,742298	6,45E-63	***

Realizado por: Jijón, A. 2023

Para obtener los resultados que se muestran en la tabla, como ya se mencionó anteriormente se realizó el ingreso de las variables manualmente de una en una, si la variable es significativa para el modelo se la sigue manteniendo, caso contrario ya no se la toma en cuenta, en otras palabras se está utilizando el método paso a paso, pero se lo realiza manualmente. Entonces se obtiene que las variables: X3, X4, son variables continuas significativas y las variables: X5, X20, son nominales y también son significativas en todos sus factores y serán consideradas para el modelo. Finalmente también los interceptos, valores de α (0|1, 1|2...) resultan todos significativos.

Tabla 23-3: Predicción versus real, regresión ordinal con criterio P-valor, instituciones urbanas

N°	Y3	
	Real	Predicho
1	1	1
2	1	0
3	1	1
4	0	1
5	2	2
6	1	2
7	1	1
8	2	1
9	1	1
10	2	2
11	1	1
12	1	1
13	2	3
14	2	2
15	3	3
16	3	2
17	3	2
18	2	3
19	4	4
20	4	4

Realizado por: Jijón, A. 2023

En la tabla se muestra un resumen de 20 predicciones realizadas con el modelo obtenido con el criterio P-valor con los datos de 20 estudiantes de las instituciones educativas fiscales urbanas, se puede ver que mayormente coinciden los valores predichos con los valores reales.

Tabla 24-3: Matriz de confusión, regresión ordinal urbana, criterio P-valor

Predicción	0	1	2	3	4
0	4	5	0	0	0
1	13	27	12	0	0
2	0	8	15	10	0
3	0	0	4	14	4
4	0	0	0	2	3
5	0	0	0	0	0

Realizado por: Jijón, A. 2023

En la tabla se puede observar la denominada matriz de confusión de los datos, es decir con esta tabla podemos darnos cuenta de cuantas respuestas son correctas y cuantas están erradas. Por ejemplo, si analizamos cuando Y3 vale 0 el modelo predice cuatro veces correctamente, esto se indica en la primera celda pintada, mientras que cinco veces predice incorrectamente, es decir, en vez de predecir 0, predice cinco veces 1; de igual manera cuando Y3 es 1, se puede ver que el modelo predice correctamente veintisiete veces correctamente, la segunda celda pintada, mientras que predice veinticinco veces incorrectamente, trece veces predice 0 y doce veces predice 2. Teniendo en consideración los errores y aciertos se llega a la conclusión que con este modelo se predice correctamente el 51,63934% y se predice incorrectamente el 48,36066%.

Tabla 25-3: Variables seleccionadas con el criterio AIC, método paso a paso automático

Variable	Coficiente
X3	0,8729589
X4	4,1479034
X59	-0,7834635
X510	-0,8849578
X102	-0,6225173
X103	-0,14623
X104	0,4156901
X105	0,5354005
X182	-0,1373346
X183	0,477455
X184	1,5882298

X202	0,4171456
0 1	36,32751
1 2	39,47267
2 3	42,26176
3 4	45,46788
4 5	48,19058

AIC: 809.411

Realizado por: Jijón, A. 2023

Para determinar las variables que influyen en el rendimiento académico clasificado como factores y estos factores son ordenados, en este caso se utiliza el método basado en la clasificación AIC, este método selecciona el conjunto de variables que mejor se ajustan, predicen la variable Y3 automáticamente comparando los valores AIC para cada conjunto de variables analizadas, en este caso el programa selecciona el conjunto de variables que arrojan el menor valor de AIC (criterio de información de Akaike): 809,411; dentro de este proceso se seleccionó el método paso a paso de introducción de variables hacia adelante y hacia atrás (both). Como resultado se obtiene que las Variables independientes: X3, X4, X5, X10, X18, X20, están relacionadas con la variable dependiente Y3. Si de estas variables obtenemos su P valor resultan significativas X3, X4, X5, X20 excepto X10, X18.

Tabla 26-3: Predicción versus real, regresión ordinal con criterio AIC, instituciones urbanas

N°	Y3	Y3
	Real	Predicho
1	1	1
2	0	1
3	1	0
4	1	0
5	2	2
6	2	1
7	1	1
8	1	2
9	1	1
10	2	2
11	1	1
12	1	1
13	3	2
14	2	2
15	3	3
16	2	3
17	2	3

18	3	2
19	4	4
20	4	4

Realizado por: Jijón, A. 2023

En la tabla se muestra un resumen de 20 predicciones realizadas con el modelo obtenido con el criterio AIC con los datos de 20 estudiantes de las instituciones educativas fiscales urbanas, se puede ver que mayormente coinciden los valores predichos con los valores reales.

Tabla 27-3: Matriz de confusión, regresión ordinal urbana, criterio AIC

Predicción	0	1	2	3	4
0	7	6	0	0	0
1	9	25	11	0	0
2	1	9	16	10	0
3	0	0	4	13	4
4	0	0	0	3	3
5	0	0	0	0	0

Realizado por: Jijón, A. 2023

En la tabla se puede observar la denominada matriz de confusión de los datos, es decir con esta tabla podemos darnos cuenta de cuantas respuestas son correctas y cuantas están erradas. Por ejemplo, si analizamos cuando Y3 vale 0 el modelo predice siete veces correctamente, esto se indica en la primera celda pintada, mientras que seis veces predice incorrectamente, es decir, en vez de predecir 0, predice seis veces 1; de igual manera cuando Y3 es 1, se puede ver que el modelo predice correctamente veinticinco veces correctamente, la segunda celda pintada, mientras que predice veinte veces incorrectamente, nueve veces predice 0 y once veces predice 2. Teniendo en consideración los errores y aciertos se llega a la conclusión que con este modelo se predice correctamente el 52,45902% y se predice incorrectamente el 47.54098%.

CAPÍTULO IV

4 RESULTADOS Y DISCUSIÓN

Ahora que ya hemos obtenido las variables predictoras para cada tipo de institución y con varios métodos, entonces vamos a analizarlos y determinar cual da mejores resultados para predecir el rendimiento académico y por ende los modelos matemáticos. En el caso de la regresión lineal múltiple, tanto para instituciones educativas fiscales rurales como urbanas comparamos los R^2 ajustados obtenidos y su error estándar, eligiendo aquel conjunto de variables que mejores ajustes presenten; con ese conjunto de variables que mejores ajustes presenten se determina el modelo o ecuación matemática, luego se realiza la predicción del rendimiento académico y se determina si el modelo obtenido arroja valores predichos cercanos a los valores reales, finalmente decidimos si el modelo es válido o no. Por otro lado para la regresión logística ordinal ya se realizó las predicciones, se obtuvieron los porcentajes de error, acierto, en este capítulo se analizarán los resultados de las predicciones, errores y aciertos y se determinará si los conjuntos de variables dan buenos resultados para posteriormente determinar el modelo y validarlo de ser el caso.

4.1 Modelo de regresión lineal múltiple, instituciones fiscales rurales

4.1.1 *Análisis de regresión lineal múltiple para instituciones educativas rurales*

Con el método de introducción paso a paso manual y el criterio P-valor se obtiene que las variables: X3, X4, X5, X7, X11, X15; resultan significativas en al menos uno de sus factores, es decir la ecuación matemática estaría compuesta por las variables mencionadas, teniendo en cuenta los niveles y factores que cada una de ellas tiene; con este método y conjunto de variables se obtiene un R^2 ajustado de 0,8056 y un Error estándar de 0,2373. En el caso del método paso a paso hacia adelante y hacia atrás automático, basado en el criterio AIC utilizando la función (stepAIC) se obtiene que el conjunto de variables que predicen el rendimiento académico son: X3, X4, X5, X7, X11, X13; este método y conjunto de variables arrojan un Error estándar de 0,2373 y R^2 de 0,8057. Se puede observar que casi todas las variables se repiten en los 2 casos, la única variable que cambia es X13 en un caso y X11 en otro; ahora si

comparamos el Error estándar y el R^2 ajustado se puede ver que son similares, en el caso del criterio P-valor el R^2 ajustado es mejor por una diezmilésima unidad. En este caso se puede elegir cualquiera de los 2 conjuntos de variables y nosotros nos quedamos con el conjunto de variables con el criterio del P-valor.

4.1.2 *Predicción del rendimiento académico, instituciones educativas fiscales rurales*

Para la predicción del rendimiento académico de los estudiantes de instituciones educativas fiscales rurales se utilizan los datos de 85 estudiantes que al momento de recolectar la información cursaban el segundo año de bachillerto, a continuación se muestran algunas de ellas.

Tabla 1-4: Valores reales versus valores predichos, utilizando el modelo de regresión lineal múltiple en Rstudio, instituciones rurales

Nº	Y2 predicho	Y2 real
1	7,471946	7,57
2	7,528221	7,29
3	7,657510	7,41
4	7,506167	7,45
5	8,088042	7,67
6	8,100682	8,25
7	8,392397	8,30
8	8,187277	8,23
9	9,030857	9,05
10	8,160608	8,54

Realizado por: Jijón, A. 2023

En la tabla se puede observar el valor predicho y el valor real del promedio de primero de bachillerato general unificado para 10 estudiantes tomados al azar, para la predicción se utiliza el comando “predict” y se lo hace sobre la base del modelo que se obtuvo. Se puede decir que son valores muy cercanos a los reales.

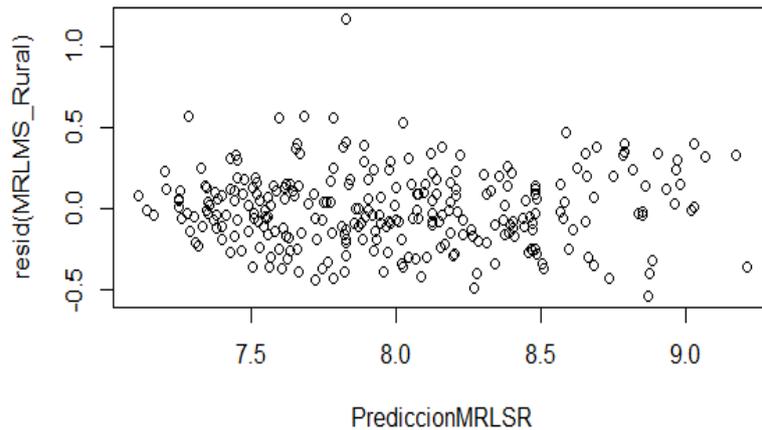


Figura 1-4: Gráfica de residuos versus valores estimados de modelo lineal rural.
 Realizado por: Jijón, A. 2023

Luego de analizar la gráfica se puede decir que los residuos están bien distribuidos con respecto a la aleatoriedad.

4.1.3 Modelo de regresión lineal múltiple para predecir el rendimiento académico de estudiantes de instituciones educativas fiscales rurales.

Una vez realizado el análisis de todas las variables propuestas, se obtiene el conjunto de variables que se involucran en la predicción, se analiza la predicción y se ve que esta arroja resultados que son aceptables, es decir se acercan mucho a la realidad, de tal manera que la ecuación queda de la siguiente manera:

$$\begin{aligned}
 Y_{2,rural} = & 1,1909195 + 0,2726728X3 + 0,5742237X4 - 0,2058176X52 \\
 & - 0,0639365X53 + 0,1088936X54 - 0,0588109X55 - 0,1873119X56 \\
 & - 0,1811521X57 - 0,0801192X72 + 0,018687X112 \\
 & - 0,0028814X113 + 0,1308653X114 + 0,0215741X152 \\
 & - 0,0001069X153 + 0,2306691X154 + 0,0525584X155 + E
 \end{aligned}$$

Las variables que predicen el rendimiento académico de las instituciones educativas fiscales rurales, mediante regresión lineal múltiple son: **X3** promedio de promoción noveno, **X4** promedio de promoción décimo, **X5** institución educativa, **X7** género, **X11** título que los representantes aspiran del estudiante, **X15** ingreso promedio mensual. Cada una de estas

variables con su respectivo coeficiente para cada uno de sus factores, por ejemplo: la variable X5, con sus factores 2, 3, 4, 5, 6, 7; el factor 1 no aparece por que el programa lo toma como referencia; entonces en el modelo aparece la variable X52, esta significa la variable X5 con su factor 2, R-studio nos arroja coeficientes para la variable en cada uno de sus factores.

4.1.4 Validación del modelo de regresión lineal múltiple para instituciones educativas fiscales rurales.

Una vez realizada la predicción, donde se cuenta con los valores reales y los valores predichos (estimados) se realiza el cálculo de los siguientes indicadores.

$$SCR = \sum_i (\hat{y}_i - \bar{y})^2 = 15,3459541$$

$$SCE = \sum_i (y_i - \hat{y}_i)^2 = 3,776663075$$

$$SCT = SCE + SCR = 19,12261718$$

$$R^2 = \frac{SCR}{SCT} = 0,802502814$$

$$S = \sqrt{\frac{SCE}{n - k - 1}} = 0,220042626$$

Dónde:

SCR: suma cuadrática de regresión

SCE: suma cuadrática del error

SCT: suma cuadrática total

R²: r cuadrado ajustado

S: error estándar

n: número de elementos para el análisis

k: número de variables en el modelo

\hat{y}_i : valores estimados o predichos

\bar{y} : promedio de valores reales

y_i: valores reales

Para realizar el cálculo se utilizan los valores reales y los valores predichos como se mencionó anteriormente de 85 estudiantes de las instituciones educativas fiscales rurales. Ahora

comparando el R cuadrado ajustado obtenido directamente por la aplicación de la regresión lineal múltiple, valores de entrenamiento, es de 0,8056 y el R cuadrado ajustado utilizando el cálculo con los valores predichos, valores de prueba, se obtiene que es 0,802502814, es muy cercano; comparando también el error estándar obtenido por la regresión en R-estudio es 0,2373 y el obtenido con el cálculo utilizando los valores predichos es 0,22460255, de igual manera un valor muy cercano; por lo tanto como se obtiene valores similares de estos indicadores mencionados, se concluye que el modelo es válido.

4.2 Modelo de regresión lineal múltiple, instituciones fiscales urbanas

4.2.1 Análisis de regresión lineal múltiple para instituciones educativas urbanas

Con el método de introducción paso a paso manual y el criterio P-valor se obtiene que las variables: X3, X4, X5, X13, X20; resultan significativas en al menos uno de sus factores, es decir la ecuación matemática estaría compuesta por las variables mencionadas, teniendo en cuenta los niveles y factores que cada una de ellas tiene; con este método y conjunto de variables se obtiene un R^2 ajustado de 0,7475 y un Error estándar de 0,303. En el caso del método paso a paso hacia adelante y hacia atrás automático, basado en el criterio AIC utilizando la función (stepAIC) se obtiene que el conjunto de variables que predicen el rendimiento académico son: X3, X4, X5, X7, X10, X13, X20; este método y conjunto de variables arrojan un Error estándar de 0,3 y R^2 de 0,7525. Se puede observar que casi todas las variables se repiten en los 2 casos, las únicas variables que incrementan con el criterio AIC es X10 y X7; ahora si comparamos el Error estándar y el R^2 ajustado se puede ver que en los dos casos mejora utilizando el criterio AIC, entonces nos quedamos con el conjunto de variables que da como resultado utilizando la función (stepAIC).

4.2.2 Predicción del rendimiento académico, instituciones educativas fiscales urbanas

Para la predicción del rendimiento académico de los estudiantes de instituciones educativas fiscales urbanas se utilizan los datos de 122 estudiantes que al momento de recolectar la información cursaban el segundo año de bachillerato, a continuación se muestran algunas de ellas.

Tabla 2-4: Valores reales versus valores predichos, utilizando el modelo de regresión lineal múltiple en Rstudio, instituciones urbanas

N°	Y2 predicho	Y2 real
1	8,093697	7,77
2	7,816100	7,89
3	7,672594	7,82
4	8,519247	8,46
5	7,499260	7,69
6	7,801720	8,38
7	8,603096	8,68
8	8,872267	8,09
9	8,512127	8,64
10	9,210338	9,58

Realizado por: Jijón, A. 2023

En la tabla se puede observar el valor predicho y el valor real del promedio de primero de bachillerato general unificado de instituciones educativas del sector urbano, para 10 estudiantes tomados al azar, para la predicción se utiliza el comando “predict” y se lo hace sobre la base del modelo que se obtuvo. Se puede decir que son valores cercanos a los reales.

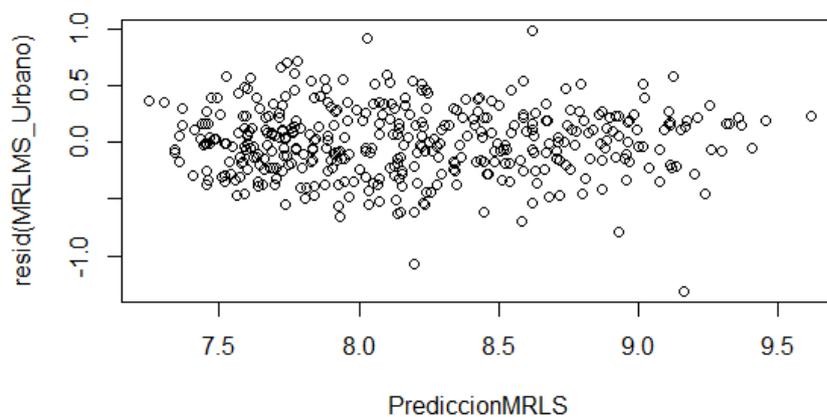


Figura 2-4: Gráfica de residuos versus valores ajustados de modelo lineal urbano.

Realizado por: Jijón, A. 2023

Luego de analizar la gráfica se puede decir que los residuos están bien distribuidos con respecto a la aleatoriedad.

4.2.3 *Modelo de regresión lineal múltiple para predecir el rendimiento académico de estudiantes de instituciones educativas fiscales urbanas.*

Una vez realizado el análisis de todas las variables propuestas, se obtiene el conjunto de variables que se involucran en la predicción, se analiza la predicción y se ve que esta arroja resultados que son aceptables, es decir se acercan mucho a la realidad, de tal manera que la ecuación queda de la siguiente manera:

$$\begin{aligned}
 Y_{2\text{urbano}} = & 1,62257 + 0,10625X3 + 0,70709X4 - 0,10434X59 - 0,11493X510 \\
 & - 0,04926X72 - 0,07328X102 - 0,02734X103 + 0,05037X104 \\
 & + 0,10788X105 + 0,22628X132 + 0,08009X202 + E
 \end{aligned}$$

Este es el modelo matemático que predice el rendimiento escolar (promedio de promoción de primero de bachillerato) para las instituciones educativas fiscales urbanas del cantón Salcedo. Si el coeficiente es positivo este indica que la variable es directamente proporcional al rendimiento académico y si el coeficiente es negativo, entonces es inversamente proporcional. Es importante mencionar que se comprobó que R-studio utiliza todos los coeficientes de todas las variables con sus respectivos factores para el modelo, incluso aquellas variables que en sus factores no fueron significativos.

Las variables que predicen el rendimiento académico de las instituciones educativas fiscales rurales, mediante regresión lineal múltiple son: **X3** promedio de promoción noveno, **X4** promedio de promoción décimo, **X5** institución educativa, **X7** género, **X10** escolaridad representante (mujer), **X13** estado civil estudiante, **X20** tiene servicio de internet. De igual manera cada variable tiene su coeficiente para cada uno de sus factores, por ejemplo: la variable X10, con sus factores 2, 3, 4, 5, es decir, X102, lo que significa la variable X10 con su factor 2, ahora no aparece el código X101, pues el factor 1 es considerado como referencia, cada variable con su respectivo factor tiene su coeficiente.

4.2.4 Validación del modelo de regresión lineal múltiple para instituciones educativas fiscales urbanas.

Una vez realizada la predicción, donde se cuenta con los valores reales y los valores predichos (estimados) se realiza el cálculo de los siguientes indicadores.

$$SCR = \sum_i (\hat{y}_i - \bar{y})^2 = 29,274010$$

$$SCE = \sum_i (y_i - \hat{y}_i)^2 = 8,674640$$

$$SCT = SCE + SCR = 37,948650$$

$$R^2 = \frac{SCR}{SCT} = 0,771411$$

$$S = \sqrt{\frac{SCE}{n - k - 1}} = 0,275850$$

Dónde:

SCR: suma cuadrática de regresión

SCE: suma cuadrática del error

SCT: suma cuadrática total

R²: r cuadrado ajustado

S: error estándar

n: número de elementos para el análisis

k: número de variables en el modelo

ŷ_i: valores estimados o predichos

ȳ: promedio de valores reales

y_i: valores reales

Para realizar el cálculo se utilizan los valores reales y los valores predichos como se mencionó anteriormente de 120 estudiantes de las instituciones educativas fiscales rurales. Ahora comparando el R cuadrado ajustado obtenido directamente por la aplicación de la regresión lineal múltiple, valores de entrenamiento, es de 0,7525 y el R cuadrado ajustado utilizando el cálculo con los valores predichos, valores de prueba, se obtiene que es 0,771411, es muy cercano; comparando también el error estándar obtenido por la regresión en R-estudio es 0,3 y el obtenido con el cálculo utilizando los valores predichos es 0,275850, de igual manera un valor muy cercano; por lo tanto como se obtiene valores similares de estos indicadores mencionados, se concluye que el modelo es válido.

4.2.5 Comparación entre Modelos de regresión lineal Múltiple obtenidos

Para el modelo de regresión lineal múltiple de las instituciones educativas fiscales urbanas se obtienen un total de 7 variables independientes más el intercepto que se consideran predictores del rendimiento escolar, mientras que para las instituciones educativas fiscales rurales se involucran en el modelo 6 variables predictoras más el intercepto. El intercepto en el caso de las instituciones educativas rurales es 1,1909195, mientras que para las instituciones educativas

urbanas es 1,62257, lo que indica que la constante del modelo es mayor para las instituciones educativas urbanas.

Las variables que se repiten o predicen para los 2 modelos son: X3 (promedio de promoción noveno año), X4 (promedio promoción décimo año), X5 (nombre de Institución Educativa a la que pertenece el estudiante), X7 (Género); X3 para las instituciones educativas urbanas tiene un coeficiente de 0,10625, mientras que para las rurales es 0,2726728, indica que para el caso de las instituciones rurales es mayor, en los 2 casos tienen un coeficiente positivo, es decir si mayor es el promedio de promoción en noveno, mayor será el promedio de promoción predicho para primero de bachillerato; X4 para las instituciones urbanas tienen un coeficiente de 0,70709, mientras que para las rurales es 0,5742237, entonces en el caso de las instituciones educativas urbanas es mayor, aporta más a la predicción de igual manera en los 2 casos el coeficiente es positivo, lo que significa que a mayor sea el promedio de promoción de décimo mayor será el promedio de promoción predicho para primero de bachillerato; X5: las instituciones educativas fiscales rurales son en total 7, mientras que las instituciones educativas rurales son en total 3 y cada institución tiene su coeficiente ya que R-Studio determina un coeficiente para cada factor que compone la variable, por ejemplo: X52 la Unidad Educativa Emilio Terán pertenece las instituciones educativas fiscales rurales y su coeficiente es $-0,2058176$; X7 (Género) en el caso de las instituciones educativas fiscales urbanas el coeficiente es $-0,04926$ (X72: indica que el estudiante es de género masculino) mientras que para las instituciones rurales el coeficiente es $-0,0801192$, en los dos casos no aportan mucho al promedio de promoción de primero de bachillerato y en los 2 casos los coeficientes son negativos, lo que indica que ser hombre disminuye el promedio.

Las variables que difieren en los modelos; X10 (escolaridad representante mujer), X13 (estado civil del estudiante), X20 (cuenta con servicio de internet), en el caso de las instituciones educativas fiscales urbanas son variables predictoras, mientras que no lo son para las instituciones educativas fiscales rurales; Las variables: X11 (título que los representantes aspiran del estudiante), X15 (ingreso promedio mensual), predicen el rendimiento académico en las instituciones educativas fiscales rurales, mientras que no lo son para las instituciones educativas fiscales urbanas.

4.3 Regresión logística ordinal, instituciones fiscales rurales

4.3.1 Análisis regresión logística ordinal, instituciones fiscales rurales

Como se mencionó en el capítulo anterior, antes de introducir los datos al programa se realizó el cambio de variable continua a variable nominal ordenada, utilizando intervalos, al primer intervalo se le asignó el valor de 0 a los promedios de promoción de primero de bachillerato cuya nota tomaba valores de 7 a 7,50; al segundo intervalo, notas mayores a 7,50 y menores o iguales a 8, etc. Luego se introducen los datos y mediante el método paso a paso manual y el método paso a paso automático con la función (stepAIC), se obtienen los siguientes resultados.

Con el criterio P-valor, es decir, paso a paso manual se obtiene que las variables: X3 promedio de promoción de noveno, X4 promedio promoción décimo, X7 género, X9 escolaridad representante (mujer), X11 título que aspiran del estudiante; son significativas o en este caso explican el rendimiento académico. Con el criterio AIC, es decir, con la función (stepAIC) se obtiene que las variables X3 promedio de promoción de noveno, X4 promedio promoción décimo, X7 género, X9 escolaridad representante (mujer), X12 interés de los representantes en el proceso educativo; son significativas o en este caso explican el rendimiento académico de las instituciones educativas fiscales rurales del cantón Salcedo.

4.3.2 Matriz de confusión, instituciones rurales

Con el conjunto de variables que resultan del análisis de regresión logística para cada método, se realiza la predicción con los datos de 84 estudiantes que conforman las instituciones educativas fiscales rurales y con esos resultados de predicción se realiza una matriz de confusión, la cual nos resume los errores y aciertos de predicción, es decir, si el conjunto de variables acierta o erra al momento de asignarle un intervalo en este caso de calificación.

En el caso del criterio P-valor, como resultado se obtienen 53 predicciones correctas y 31 predicciones erradas, lo que representa un 63,09524% correctamente y 36,90476% errado. En el caso del criterio AIC, como resultado se obtienen 53 predicciones correctas y 31 incorrectas, esto representa que el 63,09524% es correcto y el 36,90476% es errado.

En los dos criterios los resultados son iguales. Sin embargo el porcentaje de aciertos es demasiado bajo, concluyendo que los conjuntos de variables dan predicciones no confiables.

4.4 Regresión logística ordinal, instituciones fiscales urbanas

4.4.1 Análisis regresión logística ordinal, instituciones fiscales urbanas

Como se mencionó en el capítulo anterior, antes de introducir los datos al programa se realizó el cambio de variable continua a variable nominal ordenada, utilizando intervalos, al primer intervalo se le asignó el valor de 0 a los promedios de promoción de primero de bachillerato cuya nota tomaba valores de 7 a 7,50; al segundo intervalo, notas mayores a 7,50 y menores o iguales a 8, etc. Luego se introducen los datos y mediante el método paso a paso manual y el método paso a paso automático con la función (stepAIC), se obtienen los siguientes resultados.

Con el criterio P-valor, es decir, paso a paso manual se obtiene que las variables: X3 promedio de promoción de noveno, X4 promedio promoción décimo, X5 institución educativa, X20 si el estudiante cuenta con servicio de internet; son significativas o en este caso explican el rendimiento académico. Con el criterio AIC, es decir, con la función (stepAIC) se obtiene que las variables X3 promedio de promoción de noveno, X4 promedio promoción décimo, X5 institución educativa, X10 escolaridad representante (mujer), X18 tiempo para llegar a la institución, X20 si el estudiante cuenta con servicio de internet; son significativas o en este caso explican el rendimiento académico de las instituciones educativas fiscales urbanas del cantón Salcedo.

4.4.2 Matriz de confusión, instituciones urbanas

Con el conjunto de variables que resultan del análisis de regresión logística para cada método, se realiza la predicción con los datos de 121 estudiantes que conforman las instituciones educativas fiscales urbanas y con esos resultados de predicción se realiza una matriz de confusión, la cual nos resume los errores y aciertos de predicción, es decir, si el conjunto de variables acierta o erra al momento de asignarle un intervalo en este caso de calificación.

En el caso del criterio P-valor, como resultado se obtienen 63 predicciones correctas y 59 predicciones erradas, lo que representa un 51,63934% correctamente y 48,36066% errado. En el caso del criterio AIC, como resultado se obtienen 64 predicciones correctas y 58 incorrectas, esto representa que el 52,45902% es correcto y el 47,54098% es errado.

En los dos criterios los resultados son similares, en el caso del criterio AIC se dan mayores aciertos. Sin embargo el porcentaje de aciertos es demasiado bajo, concluyendo que los conjuntos de variables dan predicciones no confiables.

4.5 Planteamiento de la Hipótesis.

La Inteligencia Artificial mediante Regresión Lineal Múltiple y Regresión Logística permiten predecir el Rendimiento Académico de los estudiantes de bachillerato de las unidades educativas fiscales del Distrito Educativo 05D06 (Salcedo-Ecuador).

Hipótesis Nula: H_{01} “La Inteligencia Artificial mediante Regresión Lineal Múltiple no permite predecir el Rendimiento Académico de los estudiantes de bachillerato de las unidades educativas fiscales del Distrito Educativo 05D06 (Salcedo-Ecuador)”

Hipótesis Alternativa: H_{11} “La Inteligencia Artificial mediante Regresión Lineal Múltiple permite predecir el Rendimiento Académico de los estudiantes de bachillerato de las unidades educativas fiscales del Distrito Educativo 05D06 (Salcedo-Ecuador)”

Hipótesis Nula: H_{02} “La Inteligencia Artificial mediante Regresión Logística no permite predecir el Rendimiento Académico de los estudiantes de bachillerato de las unidades educativas fiscales del Distrito Educativo 05D06 (Salcedo-Ecuador)”

Hipótesis Alternativa: H_{12} “La Inteligencia Artificial mediante Regresión Logística permite predecir el Rendimiento Académico de los estudiantes de bachillerato de las unidades educativas fiscales del Distrito Educativo 05D06 (Salcedo-Ecuador)”

4.6 Prueba de Hipótesis

Luego de analizar todas las posibles combinaciones de variables que se relacionan con el rendimiento académico, tanto para regresión lineal múltiple y regresión logística ordinal, para instituciones educativas fiscales urbanas y rurales; determinar que el modelo de regresión lineal múltiple ofrece resultados importantes y el modelo de regresión logística ordinal no da buenos resultados, se determina los modelos matemáticos. Ahora es el momento de realizar la prueba de hipótesis.

Para la regresión lineal múltiple de instituciones educativas fiscales rurales, esta prueba se la realiza mediante el P-valor, basado en el estadístico F, cuyo valor obtenido fue: estadístico F 72,75; P-value: < 2,2e-16; como el P-valor es inferior a 0,05 se rechaza la hipótesis nula ($H_0: \beta_1 = \beta_2 \dots = \beta_k = 0$), ninguna variable explica el rendimiento académico; y se acepta la hipótesis alternativa ($H_1: \text{algún } \beta_i \neq 0$) donde el modelo conformado por las variables X3, X4, X5, X7, X11, X15, explican el rendimiento académico de manera significativa. En el caso de la

regresión lineal múltiple de instituciones educativas fiscales urbanas, esta prueba se la realiza mediante el P-valor, basado en el estadístico F, cuyo valor obtenido fue: estadístico F: 111,8; P-value: $< 2,2e-16$; como el P-valor es inferior a 0,05 se rechaza la hipótesis nula ($H_0: \beta_1 = \beta_2 \dots = \beta_k = 0$), ninguna variable explica el rendimiento académico; y se acepta la hipótesis alternativa (H_1 : algún $\beta_i \neq 0$) donde el modelo conformado por las variables X3, X4, X5, X7, X10, X13, X20, explica el rendimiento académico de manera significativa.

La Inteligencia Artificial mediante Regresión Lineal Múltiple si permite predecir el Rendimiento Académico de los estudiantes de bachillerato de las unidades educativas fiscales del Distrito Educativo 05D06 (Salcedo-Ecuador).

La Inteligencia Artificial mediante Regresión Logística no permite predecir el Rendimiento Académico de los estudiantes de bachillerato de las unidades educativas fiscales del Distrito Educativo 05D06 (Salcedo-Ecuador).

CAPÍTULO V

5 PROPUESTA

Una vez realizados los análisis llegamos a la conclusión que los modelos de regresión logística no son buenos predictores, por lo contrario los modelos de regresión lineal múltiple si son buenos predictores del rendimiento académico de los estudiantes de bachillerato de las unidades educativas fiscales del cantón Salcedo, en este capítulo proponemos los modelos matemáticos obtenidos y los recursos necesarios en caso de implementarlos, es importante mencionar que se puede implementar solamente en instituciones educativas que formaron parte de la investigación.

5.1 Justificación

Las instituciones educativas fiscales del cantón Salcedo manejan gran cantidad de información, pero yo la considero, muerta, pues es utilizada solo cuando se presenta un problema, por ejemplo, un estudiante no asiste durante 3 días a la institución educativa, recién en ese instante tanto docente tutor, personal del DECE y autoridades, empiezan a buscar información del estudiante para poder identificarlo, cuando a veces ya es muy tarde y dicho estudiante o sus representantes ya decidieron retirarle de la institución educativa, nosotros proponemos darle vida a esa información, con los modelos matemáticos obtenidos podemos programarlo mediante software en una computadora y predecir si un estudiante va a tener éxito o no durante su etapa como estudiante en bachillerato, si la predicción de un estudiante arroja resultados negativos, es decir, bajo rendimiento académico, todos los autores de la comunidad educativa: representantes, estudiantes, docentes, personal del DECE, autoridades institucionales, autoridades distritales de educación, podemos tomar acción antes que inicie el año lectivo y durante el mismo para ayudarlo y darle seguimiento al desarrollo positivo del estudiante en sus actividades académicas.

5.2 Modelos matemáticos propuestos

5.2.1 *Modelo de regresión lineal múltiple instituciones educativas fiscales urbanas.*

$$\begin{aligned}
Y2_{urbano} = & 1,62257 + 0,10625X3 + 0,70709X4 - 0,10434X59 - 0,11493X510 \\
& - 0,04926X72 - 0,07328X102 - 0,02734X103 + 0,05037X104 \\
& + 0,10788X105 + 0,22628X132 + 0,08009X202
\end{aligned}$$

5.2.2 *Modelo de regresión lineal múltiple instituciones educativas fiscales rurales.*

$$\begin{aligned}
Y2_{rural} = & 1,1909195 + 0,2726728X3 + 0,5742237X4 - 0,2058176X52 \\
& - 0,0639365X53 + 0,1088936X54 - 0,0588109X55 - 0,1873119X56 \\
& - 0,1811521X57 - 0,0801192X72 + 0,018687X112 \\
& - 0,0028814X113 + 0,1308653X114 + 0,0215741X152 \\
& - 0,0001069X153 + 0,2306691X154 + 0,0525584X155
\end{aligned}$$

5.3 Software a ser utilizado

Dado que el modelo matemático obtenido es una ecuación, y en este caso son ecuaciones lineales, la programación no requiere de un software matemático especializado, necesitamos solamente operaciones matemáticas básicas de multiplicación, sumas; además se requiere del uso de la función IF, ya que el modelo tiene variables dummy, es decir, variables que pueden tomar solamente valores de cero o uno y Excel si dispone de esta función; otro punto sumamente importante es que todas las instituciones educativas disponen de computadoras con el software Excel requerido para la implementación.

5.4 Recursos necesarios

5.4.1 *Recursos materiales y tecnológicos*

Se necesitan solamente un computadora en pleno funcionamiento en la cual pueda ser instalado el programa Excel, no se requiere de conexión a internet.

5.4.2 Recursos humanos

Para programar la ecuación matemática obtenida se requiere de una persona que tenga conocimientos medios en Excel, sobre el uso y aplicación de funciones IF; además de las operaciones de suma y multiplicación; la persona que realice la programación debería programar con formularios para dar una mejor interface con el usuario, de esta manera se puede evitar el cambio de programación por error al momento de ingresar los datos para predicción. Una vez realizada la programación teniendo en cuenta lo anteriormente mencionado, se necesita una persona que maneje el programa, para el manejo lo puede hacer cualquier persona que haya utilizado computadora ya que solamente tiene que abrir el programa dando doble click, luego la interface con el usuario solamente le irá pidiendo que seleccione la opción y que ingrese la información necesaria, en caso de ser información errónea, el programa le indicará que ingrese la información solicitada, según como lo haya realizado el programador.

5.4.3 Recurso tiempo

Para la programación del modelo matemático, si lo realiza una persona con conocimiento medio en el manejo de Excel, solamente requiere de máximo una hora y luego ya se puede hacer uso del mismo para predecir el rendimiento académico de estudiantes de bachillerato, es importante mencionar que este modelo solamente predice el rendimiento académico de estudiantes que pertenecen a las instituciones educativas que formaron parte de la investigación.

CONCLUSIONES

- La regresión lineal múltiple es una buena herramienta para determinar un modelo matemático que prediga el rendimiento académico o también para identificar las variables que influyen en el aprovechamiento de los estudiantes en las instituciones educativas, pues se han realizado varias investigaciones utilizando esta técnica y se respalda lo mencionado con los resultados de esta investigación.
- La regresión logística para el caso de las instituciones educativas fiscales del cantón Salcedo, no resultó una técnica adecuada para predecir el rendimiento académico; ya que nos da un modelo cuyas predicciones no son muy satisfactorias, es decir, comete muchos errores en la predicción del rendimiento académico.
- Las variables que influyen o predicen el rendimiento académico de los estudiantes de bachillerato en las instituciones educativas fiscales rurales son: Promedio de promoción de noveno EGB, Promedio de promoción de décimo EGB, la institución educativa a la que pertenece, género del estudiante, título que aspiran los representantes que obtenga el estudiante, ingreso económico mensual promedio.
- Las variables que influyen o predicen el rendimiento académico de los estudiantes de bachillerato en las instituciones educativas fiscales urbanas son: Promedio de promoción de noveno EGB, Promedio de promoción de décimo EGB, la institución educativa a la que pertenece, género del estudiante, título obtenido por la representante del estudiante, estado civil del estudiante, cuenta o no con servicio de internet el estudiante.

RECOMENDACIONES

- A más de utilizar el criterio del P-valor para elegir el mejor modelo, se recomienda también utilizar el método paso a paso automático, con el criterio AIC, ya que en este estudio se obtuvieron mejores resultados, reduciendo el error estándar, incrementando el r cuadrado ajustado, incluyendo variables que no resultaban significativas, mejorando la predicción.
- Una vez determinadas aquellas variables influyentes en el rendimiento académico se recomienda a las instituciones educativas, autoridades distritales, gubernamentales, establecer un plan de trabajo y ejecutarlo con el objetivo de mejorar las condiciones y mejorar el rendimiento académico de los estudiantes.
- Se recomienda realizar un estudio similar a este, considerando aquellas variables psicosociales, pedagógicas e institucionales, ya que al conocer los factores que influyen se pueden realizar acciones para favorecer el rendimiento académico de los estudiantes.

GLOSARIO

A

Académica

Perteneiente o relativo a centros oficiales de enseñanza, especialmente a los superiores · 2

Algoritmo

Un algoritmo es un procedimiento paso a paso para conseguir un fin. Es un concepto comúnmente relacionado con el ámbito de la informática · 7

Autónoma

Que trabaja por cuenta propia · 7

B

Bachillerato

Estudios de enseñanza secundaria que preceden a los superiores · 2

C

Criterio AIC

Criterio basado en el índice de Akaike para seleccionar modelos matemáticos estadísticos · 36

D

Deserción estudiantil

Es un término comúnmente utilizado en todo el mundo para referirse al abandono escolar. Se trata de aquella situación en la que el alumno después de un proceso acumulativo de separación o retiro, comienza a retirarse antes de la hora establecida por el sistema educativo sin obtener una nota o un certificado escolar · 1

E

Error estándar

El error estándar de la estimación es una forma de medir la precisión de las predicciones realizadas por un modelo de regresión · 28

Escolaridad

Escolaridad es el lapso de tiempo en el que un niño u joven acude de forma casi ininterrumpida al colegio o escuela para adquirir valores y conocimientos básicos · 18

H

Hipótesis nula

Es un punto de partida para la investigación que no se rechaza a menos que los datos de la muestra parezcan evidenciar que es falsa (es decir, que sí hay una relación entre los parámetros o fenómenos). Técnicamente, la hipótesis nula es una aplicación a la estadística del método de reducción al absurdo · 26

I

Instituciones fiscales

Instituciones provistas por el Gobierno para administrar todo lo referente a la economía interna de la nación y que busca el equilibrio en cuanto al pago de la renta y demás servicios económicos del país · 23

Inteligencia Artificial

en el contexto de las ciencias de la computación, es una disciplina y un conjunto de capacidades cognitivas e intelectuales expresadas por sistemas informáticos o combinaciones de algoritmos cuyo propósito es la creación de máquinas que imiten la inteligencia humana · 2

M

Matriz de confusión

La matriz de confusión es una matriz donde se muestra de manera ordenada los aciertos y errores que un modelo estadístico arroja en sus predicciones · 34

modelo matemático

Un modelo matemático es una representación simplificada, a través de ecuaciones, funciones o fórmulas matemáticas, de un fenómeno o de la relación entre dos o más variables · 4

P

Promedio promoción

Es el promedio calculado entre todas las notas de todas las materias que un estudiante ha obtenido durante el año lectivo · 17

Psicosocial

Psicosocial se refiere a la conducta humana y su inserción en la sociedad, el accionar individual analizado desde los aspectos sociales · 2

R

R-cuadrado ajustado

conocido también como coeficiente de determinación ajustado se utiliza en la regresión múltiple para ver el grado de intensidad o efectividad que tienen las variables independientes en explicar la variable dependiente · 26

Regresión Lineal Múltiple

Es un método estadístico que estima el resultado de una variable de respuesta utilizando dos o más variables independientes · 3

Regresión logística ordinal

es un tipo de análisis de regresión que se utiliza para predecir una variable ordinal, es decir, una variable cuyo valor existe en una escala arbitraria donde solo el orden relativo entre diferentes valores es significativo · 20

Rendimiento estudiantil

es un término utilizado para indicar un logro obtenido de los resultados de aprendizaje de una persona dentro de un cierto período de tiempo en forma de comprensión, aplicación, poder analítico y evaluación, expresado en forma de calificaciones, números o símbolos a través de evaluaciones realizadas directamente por el profesor o utilizando pruebas estandarizadas · 1

Rural

Pertenece o relativo a la vida del campo y a sus labores · 4

S

Sociodemográfico

Refiere a las características generales y al tamaño de un grupo poblacional. Estos rasgos dan forma a la identidad de los integrantes de esta agrupación, por ejemplo edad, el género, la actividad laboral y el nivel de ingresos, etc. · 2

Socioeconómico

Pertenece o relativo a los factores sociales y económicos · 2

U

Unión libre

Vínculo entre dos personas adultas o menores emancipados, que conviven como pareja monogámica y estable · 19

Urbana

Es un adjetivo que se usa para indicar algo que es perteneciente o relativo a la ciudad · 4

V

Variable dependiente

es una variable cuyo valor depende del valor de otra variable. Es decir, el valor de una variable dependiente variará en función del valor que tome la variable independiente · 20

Variables dummy

Son conocidas también como artificiales, pueden tomar valores de cero y uno, son mutuamente excluyentes · 24

Variables independientes

son las que cambian o se controlan para estudiar sus efectos en la variable dependiente · 20

Variables significativas

Son aquellas variables que resultan tener impacto o que tienen un aporte importante para predecir la variable dependiente · 33

BIBLIOGRAFÍA

- Baños, R., Fonseca, M., & Álvarez, M. (2019). Análisis de regresión lineal múltiple con SPSS: un ejemplo práctico. . *REIRE Revista d Innovación i Recerca En Educació*, 12 (2), 1–10. <https://doi.org/10.1344/reire2019.12.222704>.
- Baños, R., Ortiz, M., Baena, A., & Tristán, J. (2017). Satisfacción, motivación y rendimiento académico en estudiantes de Secundaria y Bachillerato: antecedentes, diseño, metodología y propuesta de análisis para un trabajo de investigación. *Espiral.Cuadernos del Profesorado*, 10(20), 40-50.
- Blum, W. (2003). ICMI-Study 14. Applications and modelling in mathematics education. Discussion Document. . *Special issue of ICMI – Bulletin*.
- Castrillón, Omar D., Sarache, William, & Ruiz-Herrera, Santiago. (2020). Predicción del rendimiento académico por medio de técnicas de inteligencia artificial. *Formación universitaria*, 13(1), 93-102.
- Erazo, Ó. (2018). Programa de hábitos escolares para mejorar el bajo rendimiento académico en estudiantes de bachillerato de un colegio público de Popayán-Colombia. *Encuentros vol.16 no.2 Barranquilla July/Dec*, 117 - 133. DOI: <http://dx.doi.org/10.15665/encuent.v16i02.923>.
- Garbanzo, G. M. (2014). Factores asociados al rendimiento académico tomando en cuenta el nivel socioeconómico: Estudio de regresión múltiple en estudiantes universitarios. *Revista Electrónica Educare*, Vol. 18(1) Enero-Abril., 119-154.
- García, M., Alvarado, J., & Jiménez, A. (2000). La predicción del rendimiento académico: regresión lineal versus regresión logística. *Psicothema. ol. 12, Suplem.2*, 248-252.
- Goleman, D. (2008). *Inteligencia Emocional*. USA: Peguin Random House Group USA.
- Granados, R. (2016). *Modelos de regresión lineal múltiple. Documentos de Trabajo En Economía Aplicada*, 60.
- Martínez-Otero, V., & Pérez, V. M. O. (1997). *Los adolescentes ante el estudio: causas y consecuencias del rendimiento académico* (Vol. 213). Editorial Fundamentos.
- Medina, A. (2011). *Didáctica General*. . México: Pearson Educación.

- Miranda, A. (2015). *Big Intelligence: Nuevas capacidades big data para los sistemas de vigilancia estratégica e inteligencia competitiva*. Madrid: EOI60.
- Montes, I., & Lerner, J. (2011). *Rendimiento académico de los estudiantes de pregrado de la universidad EAFIT, perspectiva cuantitativa*. Recuperado el 08 de Septiembre de 2021, de <http://www.eafit.edu.co/institucional/calidadeafit/investigacion/Documents/Rendimiento%20Academico-Perspectiva%20cuantitativa.pdf>.
- Morales, P. (2011). *Correlación y regresión, simple y múltiple. Apuntes de Facultad de Ciencias Humanas y Sociales. UP Comillas, 24*.
- Reyes, Y. (2003). *Relación entre el Rendimiento Académico, la Ansiedad ante los Exámenes, los Rasgos de Personalidad, el Autoconcepto y la Asertividad en Estudiantes del Primer Año de Psicología de la UNMSM*. Lima: Universidad Nacional Mayor de San Marcos.
- Ríos, G. (2010). *Factores sociodemográficos y rendimiento académico en la Universidad: el caso de estudiantes de abogacía de la Universidad Nacional de Córdoba* (Doctoral dissertation, Tesis inédita de doctorado). Universidad Nacional de Córdoba.
- Rodríguez López, A., Martínez Montaña, M., Vázquez Montiel, S., Cortés Riverol, J., Rosales de Gante, S., & Arévalo Ramírez, M. (2018). Factores sociodemográficos asociados al rendimiento académico en estudiantes de la licenciatura en Médico Cirujano-Partero. *Educación Médica Superior*, 32(3).
- Sánchez, E., & Ramírez, C. (2000). *Regresión Logística Binaria Múltiple. En: Regresión Logística en Salud Pública*. Granada: Escuela Andaluza de Salud Pública.
- Teigens, V. (2014). *Inteligencia Artificial General*. Stanford: Cambridge Stanford Books.
- Tironi, E., & Cavallo, A. (2014). *Comunicación Estratégica: Vivir en un mundo de señales*. Chile: Penguin Random House.
- Villarruel, R., Tapia, K., & Cárdenas, J. (2020). Determinantes del rendimiento académico de la educación media en Ecuador. *Revista Economía y Política*, núm. 32, 1-18. DOI: <https://doi.org/10.25097/rep.n32.2020.08>.

ANEXOS

ANEXO A. Código de programación transformación a variables tipo factor.

```
install.packages("car")
```

```
library(car)
```

```
urbano$Y3<- as.factor(urbano$Y3)
```

```
urbano$X5<- as.factor(urbano$X5)
```

```
urbano$X6<- as.factor(urbano$X6)
```

```
urbano$X7<- as.factor(urbano$X7)
```

```
urbano$X8<- as.factor(urbano$X8)
```

```
urbano$X9<- as.factor(urbano$X9)
```

```
urbano$X10<- as.factor(urbano$X10)
```

```
urbano$X11<- as.factor(urbano$X11)
```

```
urbano$X12<- as.factor(urbano$X12)
```

```
urbano$X13<- as.factor(urbano$X13)
```

```
urbano$X14<- as.factor(urbano$X14)
```

```
urbano$X15<- as.factor(urbano$X15)
```

```
urbano$X16<- as.factor(urbano$X16)
```

```
urbano$X17<- as.factor(urbano$X17)
```

```
urbano$X18<- as.factor(urbano$X18)
```

```
urbano$X19<- as.factor(urbano$X19)
```

```
urbano$X20<- as.factor(urbano$X20)
```

```
rural$Y3<- as.factor(rural$Y3)
```

```
rural$X5<- as.factor(rural$X5)
```

```
rural$X6<- as.factor(rural$X6)
```

```
rural$X7<- as.factor(rural$X7)
```

```
rural$X8<- as.factor(rural$X8)
rural$X9<- as.factor(rural$X9)
rural$X10<- as.factor(rural$X10)
rural$X11<- as.factor(rural$X11)
rural$X12<- as.factor(rural$X12)
rural$X13<- as.factor(rural$X13)
rural$X14<- as.factor(rural$X14)
rural$X15<- as.factor(rural$X15)
rural$X16<- as.factor(rural$X16)
rural$X17<- as.factor(rural$X17)
rural$X18<- as.factor(rural$X18)
rural$X19<- as.factor(rural$X19)
rural$X20<- as.factor(rural$X20)
rural$X21<- as.factor(rural$X21)
rural$X22<- as.factor(rural$X22)
```

ANEXO B. Código de programación regresión lineal múltiple: criterio AIC automático y criterio P-valor.

```
install.packages('MASS')
```

```
install.packages("nnet")
```

```
library(MASS)
```

```
library(nnet)
```

```
stepAIC(MRLM_Urbano<-
```

```
  lm(Y2~X2+X3+X4+X5+X7+X8+X9+X10+X11+X12+X13+X14+X15+X16+X17+X18+X19+X20, data=urbano),method=c("both"))
```

```
summary(MRLM_Urbano)
```

```
MRLMS_Urbano<-lm(Y2~
```

```
  X2+X3+X4+X5+X7+X8+X9+X10+X11+X12+X13+X14+X15+X16+X17+X18+X19+X20, data=urbano)
```

```
summary(MRLMS_Urbano)
```

```
summary(MRLM_Urbano)
```

ANEXO C. Código de programación predicción utilizando modelo de regresión lineal múltiple seleccionado.

```
MRLMS_Urbano<-lm(Y2~X3+X4+X5+X7+X10+X13+X20, data=urbano)
```

```
summary(MRLMS_Urbano)
```

```
PrediccionMRLS <-predict(MRLMS_Urbano, urbano)
```

```
PrediccionMRLS
```

ANEXO D. Código de programación gráfico de predicción versus residuos.

```
plot(resid(MRLMS_Urbano)~PrediccionMRLS)
```

ANEXO E. Código de programación Regresión Logística Ordinal criterio AIC automático y criterio P-valor.

```
install.packages('MASS')

install.packages("nnet")

library(MASS)

library(nnet)

urbano$Y3<- as.ordered(urbano$Y3)

str(urbano)

stepAIC(ORDINAL_Urbanos<- polr(Y3~X3+X4+X5+X10+X18+X20, data=urbano,
Hess=TRUE),method=c("both"))

ORDINAL_Urbano<- polr(Y3~X3+X4+X5+X20, data=urbano, Hess=TRUE)

summary(ORDINAL_Urbano)

(ctable<-coef(summary(ORDINAL_Urbano)))

pvalor7<- pnorm(abs(ctable[, "t value"]), lower.tail = FALSE)*2

(ctable<-cbind(ctable, "p value" = pvalor7))
```

ANEXO F. Código de programación Cambio de variables a variable tipo factor para predicción con regresión logística ordinal.

```
urbanoPtabla$Y3<- as.ordered(urbanoPtabla$Y3)
```

```
urbanoPtabla$X5<- as.factor(urbanoPtabla$X5)
```

```
urbanoPtabla$X10<- as.factor(urbanoPtabla$X10)
```

```
urbanoPtabla$X18<- as.factor(urbanoPtabla$X18)
```

```
urbanoPtabla$X20<- as.factor(urbanoPtabla$X20)
```

```
str(urbanoPtabla)
```

ANEXO G. Código de programación Predicción con regresión logística ordinal.

```
pred7<-predict(ORDINAL_Urbano, urbanoPtabla)
```

```
print(pred7)
```

ANEXO H. Código de programación Matriz de confusión, porcentajes de error y acierto para regresión logística ordinal.

```
(table7<-table(pred7,urbanoPtabla$Y3))
```

```
(sum(diag(table7))/sum(table7))*100
```

```
(1-sum(diag(table7))/sum(table7))*100
```



epoch

Dirección de Bibliotecas y
Recursos del Aprendizaje

UNIDAD DE PROCESOS TÉCNICOS Y ANÁLISIS BIBLIOGRÁFICO Y
DOCUMENTAL

REVISIÓN DE NORMAS TÉCNICAS, RESUMEN Y BIBLIOGRAFÍA

Fecha de entrega: 17 / 07 / 2023

INFORMACIÓN DEL AUTOR/A (S)
Nombres – Apellidos: <i>Klever Antonio Jijón Bautista</i>
INFORMACIÓN INSTITUCIONAL
<i>Instituto de Posgrado y Educación Continua</i>
Título a optar: <i>Magíster en Matemática mención Modelación y Docencia</i>
f. Analista de Biblioteca responsable: Lic. Luis Caminos Vargas Mgs.



Firmado electrónicamente por:
**LUIS ALBERTO
CAMINOS VARGAS**



0066-DBRA-UTP-IPEC-2023