

Estimación de probabilidades de deserción escolar en la provincia de Los Ríos con métodos LOGIT: Un estudio basado en datos del ENEMDU

Estimating school dropout probabilities in Los Ríos province using logit methods: A study based on data from the ENEMDU

<https://doi.org/10.5281/zenodo.17137625>

AUTOR:

Jorge Adrian Sarcos González¹

Investigador independiente

 <https://orcid.org/0009-0007-6645-1567>

jorgesarcos95@gmail.com

DIRECCIÓN PARA CORRESPONDENCIA: jorgesarcos95@gmail.com

Fecha de recepción: 17 / 01 / 2025

Fecha de aceptación: 09 / 03 / 2025

RESUMEN

Este estudio analiza los factores cuantitativos asociados a la deserción escolar en la provincia de Los Ríos, Ecuador, utilizando microdatos de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU 2024). A través de un modelo logístico binario, se estimaron las probabilidades de que un menor de entre 5 y 17 años no asista a clases, en función de variables demográficas, económicas y contextuales. Los resultados revelan que la edad y el hecho de que el menor haya trabajado recientemente incrementan de forma significativa la probabilidad de deserción escolar. Sorprendentemente, la condición de pobreza y pobreza extrema no muestran un efecto estadísticamente significativo sobre la asistencia, una vez controladas las demás variables. El modelo presenta un ajuste robusto y una excelente capacidad de clasificación, y los efectos marginales permiten identificar con precisión el peso de cada predictor. Estos hallazgos

generan evidencia empírica actualizada sobre las brechas educativas en contextos subnacionales y muestran la utilidad de herramientas cuantitativas para orientar políticas focalizadas en territorios donde la deserción escolar sigue siendo un reto urgente y persistente.

PALABRAS CLAVE: *Modelos logísticos, deserción escolar, ENEMDU, Los Ríos, Ecuador.*

ABSTRACT

This study analyzes the quantitative determinants of school dropout in Los Ríos province, Ecuador, using microdata from the 2024 National Survey on Employment, Unemployment and Underemployment (ENEMDU). A binary logistic regression model was used to estimate the probability that children and adolescents aged 5 to 17 are not attending school, based on demographic, economic, and contextual variables. The results show that age and recent work activity significantly increase the likelihood of school dropout. Surprisingly, poverty and extreme poverty do not appear as statistically significant predictors, once the other variables are controlled for. The model demonstrates robust fit and excellent classification accuracy, and marginal effects allow for precise identification of the weight of each predictor. These findings provide updated empirical evidence on educational inequalities in subnational territories, and highlight the usefulness of quantitative tools to guide targeted policies in regions where school dropout remains a persistent and urgent challenge.

KEYWORDS: *Logistic models, school dropout, ENEMDU, Los Ríos, Ecuador.*

INTRODUCCIÓN

La deserción escolar continúa siendo uno de los principales retos para el desarrollo humano en América Latina. A pesar de las mejoras en cobertura educativa, millones de niños y adolescentes abandonan el sistema educativo antes de completar su formación básica, perpetuando ciclos de pobreza y exclusión social. En su informe regional más reciente, el Banco Interamericano de Desarrollo advierte que el 27 % de los jóvenes entre 18 y 24 años en la región no culmina la secundaria, con amplias disparidades socioeconómicas: mientras en el quintil más rico la tasa de abandono es del 14 %, en el más pobre alcanza el 39 % (Inter-American Development Bank, 2024).

En el caso ecuatoriano, este fenómeno también persiste, aunque con escasa información actualizada para evaluarlo territorialmente. La pandemia de COVID-19 agravó el panorama: los cierres prolongados de escuelas y la caída del ingreso familiar empujaron a miles de adolescentes fuera del sistema educativo, especialmente en zonas rurales. Según un metaanálisis internacional, la combinación de pérdida de ingresos y confinamiento aumentó tanto las horas trabajadas por menores como el abandono escolar, con consecuencias que podrían extenderse durante generaciones (Bracco et al., 2025).

En este contexto, la provincia de Los Ríos representa un caso particularmente crítico. Altos niveles de pobreza, empleo informal y ruralidad confluyen para generar un entorno donde la permanencia escolar se ve sistemáticamente amenazada. Sin embargo, existen muy pocos estudios cuantitativos que permitan estimar con precisión cómo influyen variables como el trabajo infantil, la edad, la pobreza o el género en la probabilidad de asistir a clases en esta provincia. El presente artículo busca llenar ese vacío mediante el uso de datos recientes de la ENEMDU 2024 y la estimación de un modelo logit, con el objetivo de identificar los factores más relevantes que determinan la asistencia escolar de niños y adolescentes entre 5 y 17 años en Los Ríos.

Revisión De La Literatura

Diversos estudios han documentado que el trabajo infantil sigue siendo una de las principales amenazas para la permanencia escolar, tanto por la competencia directa entre trabajo y estudio como por los entornos de vulnerabilidad que lo rodean. Según estimaciones recientes de la OIT y UNICEF (2024), 160 millones de niños y niñas seguían trabajando a nivel global, de los cuales casi la mitad lo hacía en condiciones peligrosas. Si bien América Latina tiene una tasa promedio inferior al promedio mundial, las disparidades territoriales, étnicas y socioeconómicas siguen siendo marcadas, especialmente en zonas rurales y comunidades indígenas (International Labour Organization, 2023).

Un informe de la OIT centrado en poblaciones indígenas evidencia que la falta de servicios educativos adecuados en zonas rurales andinas aumenta significativamente la probabilidad de que los menores trabajen en lugar de asistir a clases, profundizando así un círculo vicioso de pobreza y exclusión educativa (International Labour Organization, 2023). Este fenómeno no solo tiene implicaciones inmediatas, sino que reduce las posibilidades de movilidad social a largo plazo.

La evidencia también sugiere que la edad es un determinante clave del abandono escolar. A medida que los menores se acercan a la adolescencia, aumenta la probabilidad de deserción, debido a factores como el ingreso al mercado laboral informal, el inicio de uniones tempranas o el debilitamiento del vínculo con la escuela. Un estudio logístico para más de 700 niños en Sudán encontró que cada año adicional de edad incrementa en 15 % la probabilidad de no asistir a clases, incluso después de controlar por ingreso, distancia y acceso a servicios básicos (Alhassan & Hassan, 2024). Este patrón se repite en América Latina, donde el paso de primaria a secundaria representa el punto más crítico de abandono escolar.

Más recientemente, la literatura ha incorporado determinantes relacionales y psicosociales. El modelo IAFREE, validado con más de 15 000 estudiantes en Brasil y Chile, identifica cinco dimensiones de riesgo —vínculo con la escuela, la familia, los docentes, los pares y la comunidad— que anteceden al abandono escolar, incluso antes de que se acumulen inasistencias críticas (De Vasconcelos et al., 2023). Estos factores, sin embargo, siguen poco integrados en los análisis cuantitativos de política educativa.

Adicionalmente, la pandemia de COVID-19 reconfiguró los patrones tradicionales de exclusión. Un estudio reciente en México mostró que la brecha digital fue determinante en la permanencia escolar durante la pandemia: los municipios con mayor conectividad amortiguaron la caída en la matrícula, mientras que los hogares sin internet enfrentaron tasas de abandono tres veces más altas (Binnis & Davalos, 2023). La desigualdad digital se convierte así en una nueva variable crítica.

Pese a esta abundante evidencia regional, el caso de la provincia de Los Ríos sigue poco explorado desde un enfoque cuantitativo riguroso. La mayoría de estudios sobre deserción escolar en Ecuador se concentran en el nivel nacional, sin atender las dinámicas específicas de provincias rurales con alta informalidad laboral. Este artículo contribuye a esa agenda pendiente, cuantificando el impacto del trabajo infantil y otros factores en la probabilidad de asistir a clases, utilizando datos actualizados y técnicas de elección binaria.

METODOLOGÍA

Enfoque y Diseño de la investigación

La presente investigación adopta un enfoque cuantitativo, de tipo explicativo y diseño transversal. Su objetivo es identificar los determinantes estadísticamente significativos de la asistencia escolar entre niños, niñas y adolescentes de la provincia de Los Ríos, utilizando microdatos representativos de la ENEMDU 2024 (Instituto Nacional de Estadística y Censos, 2024). El estudio se estructura a partir de un modelo econométrico de regresión logística binaria, adecuado para analizar una variable dependiente dicotómica como la asistencia a clases (sí/no). Esta elección metodológica permite cuantificar el efecto de variables individuales y del hogar —como edad, sexo, trabajo infantil, pobreza monetaria, pobreza extrema y zona de residencia— sobre la probabilidad de asistir a la escuela. Se trata, por tanto, de un análisis causal y multivariado que busca aportar evidencia empírica rigurosa sobre una problemática educativa poco estudiada con datos recientes y modelos robustos en contextos locales como Los Ríos.

Base de datos y población de estudio

La información proviene de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) 2024, publicada por el Instituto Nacional de Estadística y Censos (INEC). Se filtró la muestra para incluir únicamente a niños, niñas y adolescentes de entre 5 y 17 años residentes en la provincia de Los Ríos, que declararon vivir con al menos un adulto responsable en el hogar. Luego del filtrado y la depuración por datos faltantes, la muestra final quedó compuesta por 1377 observaciones válidas, tamaño suficiente para estimar un modelo de elección binaria con varios predictores sin pérdida considerable de potencia estadística (Instituto Nacional de Estadística y Censos, 2024). Esta aproximación empírica se alinea con otros estudios recientes que exploran fenómenos sociales complejos a partir de microdatos representativos y modelos estadísticos multivariantes (Sarcos González, 2025).

Variable dependiente

La variable de interés es la asistencia escolar, registrada en el campo p07 de la ENEMDU. Fue recodificada como un indicador binario:

$$p_{07} = \begin{cases} 1 & \text{si el menor asiste a clases} \\ 0 & \text{si el menor no asiste a clases} \end{cases}$$

Al tratarse de un resultado dicotómico, se aplicó un modelo de regresión logística (logit), cuya función de enlace logística asegura que las probabilidades pronosticadas estén siempre dentro del intervalo [0,1]. El modelo tiene la forma:

$$Pr(Y_i = 1 | X_i) = \frac{e^{X_i' \beta}}{1 + e^{X_i' \beta}}$$

donde X_i es el vector de características del individuo y β es el vector de parámetros a estimar.

Este enfoque es el método estándar en la investigación aplicada sobre abandono y retención escolar (Alhassan & Hassan, 2024).

Variables explicativas

Se seleccionaron seis predictores claves, justificados por la teoría y evidencia empírica previa:

- **Edad (p03):** medida en años completos. Diversos estudios indican que la probabilidad de abandono escolar aumenta a partir de los 12–13 años, coincidiendo con el inicio de la adolescencia y la mayor exposición a trabajo, uniones tempranas o desmotivación educativa.
- **Sexo (p02):** recodificado como variable dummy (1 = hombre, 0 = mujer), para captar posibles desigualdades de género en la permanencia escolar.
- **Relación de parentesco (p04):** variable que indica el vínculo del menor con el jefe o jefa del hogar, recodificada como binaria (1 = hijo/a o hijastro/a, 0 = otros). Se incorporó para identificar si el grado de cercanía familiar influye en la probabilidad de asistir a clases, bajo la hipótesis de que niños sin un vínculo directo podrían tener menor protección o incentivos para asistir regularmente.
- **Zona de residencia (área):** variable dummy urbana-rural, ya que la distancia a las instituciones educativas y la disponibilidad de servicios escolares suelen diferir significativamente entre zonas urbanas y rurales.
- **Trabajo infantil (p20):** indicador binario que vale 1 si el menor trabajó al menos una hora durante la semana anterior. Esta variable constituye el foco del estudio, pues la inserción laboral temprana compite directamente con la asistencia a clases.
- **Pobreza monetaria (pobreza) y pobreza extrema (epobreza):** indicadores oficiales del INEC, contruidos a partir del ingreso familiar per cápita. Su inclusión permite verificar si las restricciones económicas siguen siendo

determinantes fundamentales de la deserción, una vez controlado el trabajo infantil.

Todas las variables explicativas fueron incluidas como métricas (en el caso de la edad) o binarias (en el caso de los indicadores), sin transformaciones adicionales. Esto facilita la interpretación directa de los coeficientes y, especialmente, de los efectos marginales.

Validación de supuestos

Antes de la estimación, se evaluó la multicolinealidad entre las variables independientes mediante el Factor de Inflación de la Varianza (VIF). Este indicador mide cuánto se inflan las varianzas de los estimadores por la colinealidad entre predictores (Dormann et al., 2013).

$$VIF_j = \frac{1}{1 - R_j^2}$$

donde R_j^2 es el coeficiente de determinación de la regresión de la variable j sobre las demás.

Variable	VIF
area	1.030
p02	1.016
p03	1.079
p04	1.054
p20	1.050
pobreza	1.160
epobreza	1.154

Tabla 1. Multicolinealidad en variables predictoras

Los resultados obtenidos (rango VIF: 1.016 – 1.154) estuvieron muy por debajo del umbral crítico de 5, lo cual indica ausencia de multicolinealidad relevante y asegura estabilidad estadística del modelo (Dormann et al., 2013) .

Este paso es fundamental: si dos o más predictores captaran información redundante, sus errores estándar aumentarían y se perdería precisión para contrastar hipótesis.

Estimación del modelo

El modelo logit fue estimado mediante máxima verosimilitud, tal como lo implementa el paquete *statsmodels* en Python (Makowski et al., 2025). Cada coeficiente β_j representa

el cambio en la razón de probabilidades de que un menor asista a clases ante un cambio marginal en la covariable x_{ij} , manteniendo las demás constantes (Dey et al., 2025; Nigri et al., 2025; Niu, 2020).

$$\text{Logit}(p_i) =$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 \cdot \text{área}_i + \beta_2 \cdot \text{p02}_i + \beta_3 \cdot \text{p03}_i + \beta_4 \cdot \text{p04}_i + \beta_5 \cdot \text{p20}_i + \beta_6 \cdot \text{pobreza}_i + \beta_7 \cdot \text{epobreza}_i$$

Donde:

Variable	Descripción
p_i	Probabilidad de que el individuo asista a clases
área_i	Área de residencia (1 = urbana, 0 = rural)
p02_i	Sexo del individuo (0 = hombre, 1 = mujer)
p03_i	Edad del individuo, en años
p04_i	Relación de parentesco con el jefe del hogar (codificada numéricamente)
p20_i	Trabajó la semana pasada (1 = sí, 0 = no)
pobreza_i	Indicador de pobreza (1 = pobre, 0 = no pobre)
epobreza_i	Indicador de pobreza extrema (1 = sí, 0 = no)

Tabla 2. Descripción de variables

Dado que esta escala logarítmica no es intuitiva para usuarios no técnicos, se calcularon los efectos marginales promedio (AME, por sus siglas en inglés), definidos como:

$$AME_j = \frac{1}{n} \sum_{i=1}^n \frac{\partial \text{Pr}(Y_i = 1)}{\partial x_{ij}} = \frac{1}{n} \sum_{i=1}^n \beta_j \cdot \text{Pr}(Y_i = 1) (1 - \text{Pr}(Y_i = 1))$$

Los AME permiten traducir los resultados del modelo al lenguaje de políticas públicas: cuántos puntos porcentuales cambia la probabilidad de asistir a clases cuando varía cada variable, manteniendo todo lo demás constante. Esta práctica es ampliamente recomendada en estudios empíricos aplicados (Howell-Moroney, 2024; Makowski et al., 2025; Niu, 2020).

Evaluación del poder predictivo

Para evaluar la capacidad de discriminación del modelo se utilizó la curva ROC (Receiver Operating Characteristic) y su correspondiente AUC (Área Bajo la Curva). La curva ROC traza la tasa de verdaderos positivos contra la tasa de falsos positivos a lo largo de todos los posibles umbrales de clasificación. El AUC resume ese desempeño en un valor entre 0 y 1, donde:

- $AUC > 0.7$ se considera discriminación muy buena.
- $AUC > 0.8$ se considera discriminación excelente.

RESULTADOS

Después de delimitar la muestra, se procedió a describir las características sociodemográficas y escolares de los menores incluidos en el estudio. La edad de los individuos se encuentra entre los 5 y los 17 años, con una distribución que evidencia un mayor número de observaciones entre los 12 y 17 años, y un menor número en los grupos de 8 a 10 años. La Figura siguiente muestra esta distribución, con picos marcados en los extremos del rango etario analizado, lo cual puede deberse a patrones de reporte, cambios en el sistema escolar o sesgos en la recolección de datos.

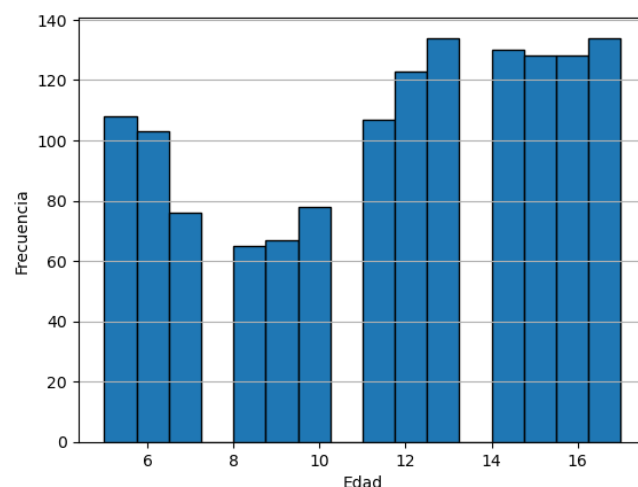


Figura 1. Distribución de edad de individuos de la muestra

En cuanto a la variable dependiente (asistencia escolar), se identificaron 112 menores que no asistían a clases al momento de la encuesta. La Figura siguiente presenta la distribución de estos casos según el sexo y el área de residencia. Se observa que los varones no asistentes son más numerosos en áreas urbanas, mientras que en zonas rurales la brecha de género es menos pronunciada.

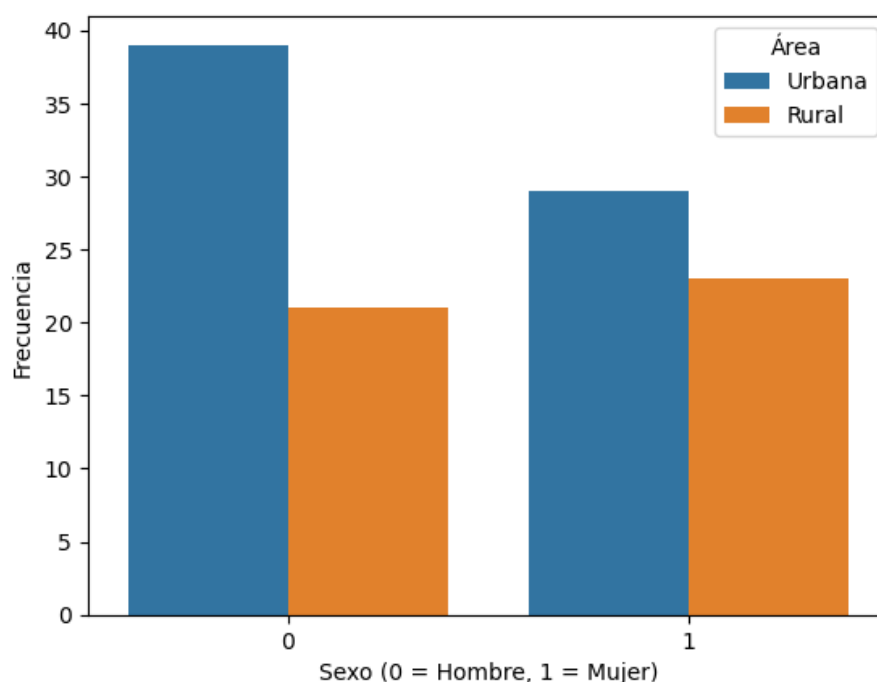


Figura 2. Distribución por sexo y área de menores que no asisten a clases

Este patrón sugiere que las dinámicas de abandono escolar podrían estar asociadas no solo al sexo del menor, sino también al entorno geográfico, lo que justifica su inclusión en el modelo de regresión. Por otro lado, se incluyeron variables explicativas que capturan condiciones individuales y del hogar. Entre ellas, la edad se considera una variable crítica, dada la evidencia empírica que sugiere que la probabilidad de deserción escolar aumenta con el paso del tiempo, especialmente a partir de los 12 o 13 años. Asimismo, el sexo (codificado como 1 = hombre, 0 = mujer) permite explorar si existen disparidades por género en la permanencia educativa. La zona de residencia (área) diferencia entre entornos urbanos y rurales, lo cual es relevante dado que las condiciones de acceso a la escuela suelen ser dispares.

Una de las variables centrales del estudio es el trabajo infantil (p20), definida como un indicador binario que toma el valor de 1 si el menor trabajó al menos una hora la semana

anterior. Dada su naturaleza estructural, se espera que tenga un impacto negativo sobre la probabilidad de asistencia escolar. Además, se incorporaron dos indicadores socioeconómicos oficiales del INEC: pobreza y pobreza extrema, contruidos a partir del ingreso familiar per cápita. Finalmente, se añadió la variable nivel de instrucción del jefe del hogar (p04), considerada como proxy del capital educativo del entorno familiar. A pesar de no presentar significancia estadística, su inclusión responde a la literatura previa que ha documentado la influencia del entorno educativo del hogar en la continuidad escolar de los menores.

Resultados de modelo logit

El modelo logit estimado sobre la asistencia escolar de niños y adolescentes en la provincia de Los Ríos presenta un ajuste robusto y hallazgos significativos. Con 1377 observaciones válidas, el modelo convergió exitosamente y obtuvo un log-verosímil de -322.25, frente a un log-verosímil nulo de -385.91. El estadístico de razón de verosimilitud (LLR) tiene un valor-p menor a 0.001, lo que indica que, en conjunto, los predictores incorporados explican significativamente la variación en la asistencia a clases.

Variable	Coef. (β)	Error Estándar	z	p-valor	Intervalo de Confianza 95 %	Significancia
Constante	4.7974	0.540	8.891	0.000	[3.740, 5.855]	***
Zona (área)	0.3078	0.225	1.371	0.170	[-0.132, 0.748]	
Sexo (p02)	-0.3419	0.224	-1.529	0.126	[-0.780, 0.096]	
Edad (p03)	-0.1773	0.036	-4.980	0.000	[-0.247, -0.108]	***
Parentesco (p04)	0.2923	0.241	1.215	0.224	[-0.179, 0.764]	
Trabajo (p20)	-3.3328	0.452	-7.378	0.000	[-4.218, -2.447]	***
Pobreza	-0.2045	0.262	-0.779	0.436	[-0.719, 0.310]	
Ext. pobreza	-0.1836	0.503	-0.365	0.715	[-1.169, 0.802]	

Tabla 3. Resultados de modelo logit

Variable	($\partial \text{Pr} / \partial x$)	Error Estándar	z	p-valor	Intervalo de Confianza 95 %	Significancia
Zona (area)	0.0190	0.014	1.366	0.172	[-0.008, 0.046]	
Hombre (p02)	-0.0211	0.014	-1.522	0.128	[-0.048, 0.006]	

Edad (p03)	-0.0109	0.002	-4.717	0.000	[-0.015, -0.006]	***
Parentesco (p04)	0.0180	0.015	1.212	0.226	[-0.011, 0.047]	
Trabajó (p20)	-0.2058	0.027	-7.664	0.000	[-0.258, -0.153]	***
Pobreza	-0.0126	0.016	-0.778	0.436	[-0.044, 0.019]	
Ext. pobreza	-0.0113	0.031	-0.365	0.715	[-0.072, 0.050]	

Tabla 4. Efectos marginales del modelo logit

Leyenda de significancia estadística

- ***** (tres asteriscos)** indican que el valor p es menor a 0.01. Esto significa que hay menos de 1 % de probabilidad de que el efecto estimado se deba al azar, por lo que se considera altamente significativo.
- **** (dos asteriscos)** se usan cuando el valor p está entre 0.01 y 0.05, indicando que hay menos de 5 % de probabilidad de error tipo I, lo cual se interpreta como significancia moderada.
- ***(un asterisco)** corresponde a valores p entre 0.05 y 0.10, señalando una significancia estadística débil, pero aún aceptable en contextos exploratorios o muestras pequeñas.
- **Si no hay asterisco** significa que el efecto de la variable es producto del azar

Una variable significativa sugiere que su efecto sobre la probabilidad de asistir a clases no es producto del azar, sino que hay evidencia estadística de una relación real, bajo los supuestos del modelo.

Interpretación de los coeficientes y efectos marginales

Dos variables aparecen como altamente significativas en el modelo ($p < 0.01$): la edad del menor y si trabajó la semana previa a la encuesta.

La edad (p03) tiene un coeficiente negativo y significativo ($\beta = -0.1773$, $p < 0.001$). Esto implica que, por cada año adicional, las probabilidades de asistir a clases se reducen. En términos más interpretables, el efecto marginal promedio (AME) es de -0.0109 , lo que significa que, en promedio, cada año adicional de edad reduce en 1.09 puntos porcentuales la probabilidad de asistir a clases, manteniendo constantes las demás variables. Este hallazgo coincide con la literatura, que identifica mayores tasas de abandono a partir de la adolescencia.

La variable más influyente es el trabajo infantil (p20). Su coeficiente β es de -3.3328 ($p < 0.001$), lo que representa una reducción drástica en los odds de asistir a clases. El efecto marginal asociado es de -0.2058 , lo que se traduce en una reducción promedio de 20.58 puntos porcentuales en la probabilidad de asistir entre quienes trabajaron al menos una hora en la semana previa, en comparación con quienes no lo hicieron. Este resultado refuerza la hipótesis central del estudio: la incorporación temprana al mercado laboral compete directamente con la permanencia escolar.

Las demás variables incluidas no alcanzaron niveles convencionales de significancia estadística ($p > 0.10$), aunque algunas presentan señales que merecen discusión:

El sexo del menor (p02) tiene un coeficiente negativo (-0.3419) y un AME de -0.0211 , lo que sugiere que los hombres podrían tener una menor probabilidad de asistir que las mujeres. Sin embargo, esta diferencia no es estadísticamente significativa ($p = 0.128$).

La variable de zona (área urbana o rural) muestra un coeficiente positivo ($\beta = 0.3078$), lo que podría indicar mejores tasas de asistencia en zonas urbanas, pero el valor $p = 0.170$ sugiere que esta diferencia no es concluyente.

Las variables de pobreza y pobreza extrema no fueron significativas en este modelo, lo cual es interesante: una vez que se controla por trabajo infantil y otras condiciones, la pobreza por ingreso per cápita no aparece como un factor determinante directo en la decisión de asistir o no a clases.

El desempeño del modelo logit se evaluó mediante la curva ROC, el área bajo la curva (AUC) y el análisis del umbral de clasificación óptimo (Iparraguirre et al., 2023). En primer lugar, la curva ROC refleja la capacidad discriminatoria del modelo para diferenciar entre menores que asisten y no asisten a clases, representando la tasa de verdaderos positivos (TPR) frente a la tasa de falsos positivos (FPR) para distintos puntos de corte. El área bajo la curva obtenida fue de 0.7584, lo cual indica un buen poder predictivo del modelo. Este valor es claramente superior al umbral de 0.5 asociado a una clasificación aleatoria, y se ubica en el rango generalmente aceptado como "bueno" en ciencias sociales, lo que respalda la calidad del ajuste.

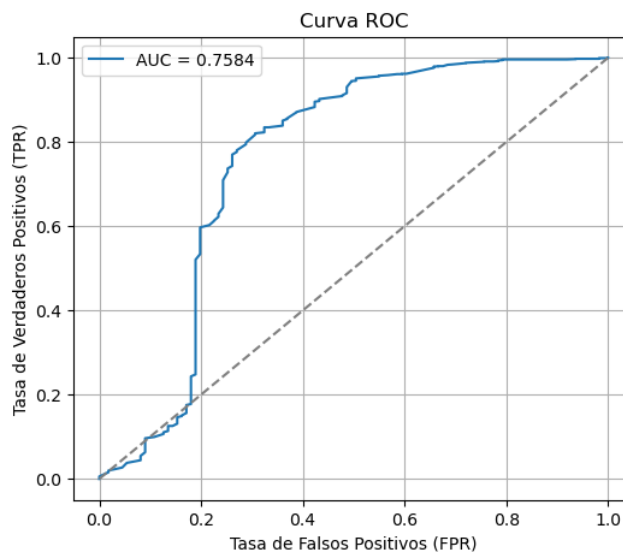


Figura 3. Curva ROC

A continuación, se empleó el estadístico de Youden ($J = \text{TPR} - \text{FPR}$) para identificar el umbral de clasificación óptimo (Qiu et al., 2023). El valor de corte que maximiza esta métrica fue de 0.90, lo que implica que el modelo solo clasifica a un menor como asistente a clases si la probabilidad estimada es igual o superior al 90 %. Este umbral conservador minimiza los falsos positivos, es decir, reduce la probabilidad de clasificar incorrectamente a un menor como asistente cuando en realidad no lo es. Esta elección de umbral resulta especialmente relevante en el contexto educativo, donde las consecuencias de una mala clasificación pueden afectar la asignación de recursos y el diseño de políticas públicas focalizadas.

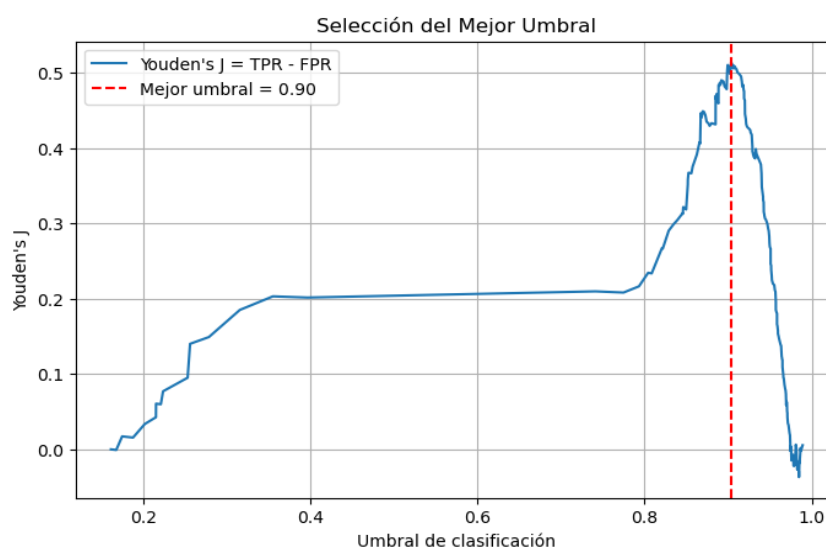


Figura 4. Selección de mejor umbral según estadístico de Youden

Finalmente, se analizó la evolución de las métricas de precisión, recall (sensibilidad) y F1-score según el umbral (Reibnegger & Schrabmair, 2014). La precisión se mantuvo elevada en la mayoría del rango, superando el 90 %, lo cual indica que entre los menores clasificados como asistentes, la mayoría efectivamente asisten. El recall, en cambio, disminuye a medida que aumenta el umbral, lo que significa que el modelo comienza a pasar por alto a algunos menores que sí asisten. Esta disminución es compensada parcialmente por el hecho de que el modelo reduce los falsos positivos. El F1-score, que combina ambas métricas, se mantiene alto hasta el umbral óptimo de 0.90 y luego cae abruptamente. Esto confirma que dicho umbral ofrece un equilibrio razonable entre precisión y sensibilidad.

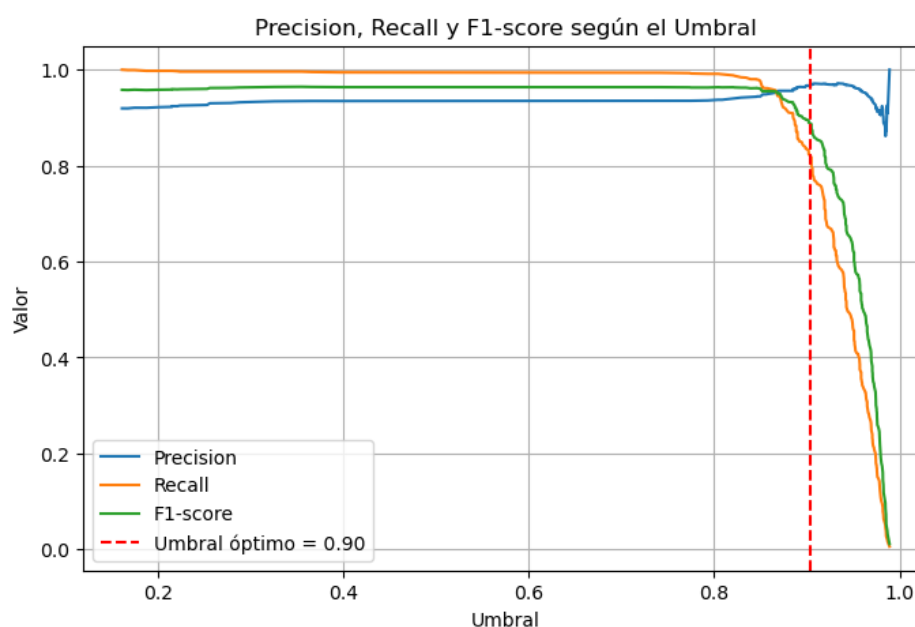


Figura 5. Métricas de precisión del modelo

Estos resultados sugieren que el modelo logit presenta un desempeño adecuado para predecir la asistencia escolar en menores de la provincia de Los Ríos.(Díaz-Coto et al., 2024). Su capacidad predictiva, unida a la elección cuidadosa del umbral, permite establecer una herramienta útil para identificar factores asociados a la deserción escolar y diseñar intervenciones basadas en evidencia.

CONCLUSIONES

Los resultados obtenidos en este estudio revelan hallazgos estadísticamente sólidos y metodológicamente robustos sobre los factores asociados a la deserción escolar en la provincia de Los Ríos. El análisis logit permitió estimar la probabilidad de que un menor entre 5 y 17 años no asista a clases en función de variables demográficas y socioeconómicas, usando microdatos representativos de la ENEMDU 2024.

Una de las principales conclusiones es la fuerte asociación negativa entre la edad y la probabilidad de asistencia escolar: conforme aumentan los años, la probabilidad de abandonar el sistema educativo también lo hace, con efectos marginales significativos a niveles convencionales. Este patrón se alinea con lo documentado en la literatura latinoamericana, donde el tránsito entre la primaria y la secundaria representa un punto crítico en la trayectoria escolar. Tal hallazgo exige la implementación de políticas específicas para adolescentes, incluyendo modalidades educativas flexibles, tutorías reforzadas y mecanismos de reenganche.

En contraste, variables tradicionalmente consideradas como determinantes estructurales —como la pobreza monetaria y la pobreza extrema— no mostraron efectos significativos sobre la probabilidad de deserción en el modelo estimado, una vez controlado el resto de factores. Este resultado es sorprendente, y sugiere que, al menos en el contexto de Los Ríos, las restricciones económicas directas podrían estar siendo compensadas por transferencias, redes familiares o la propia gratuidad del sistema educativo. No obstante, esto no debe interpretarse como irrelevancia de la pobreza en general, sino como evidencia de que sus efectos podrían estar mediados por otros factores no observados o captados en el modelo.

La ausencia de significancia estadística de los indicadores de pobreza y pobreza extrema puede entenderse si se considera que el modelo controla simultáneamente edad, residencia y participación laboral. Al medir la pobreza únicamente por ingreso per cápita, se dejan fuera dimensiones como el capital cultural, la conectividad o la cohesión social, que influyen de forma más directa en la trayectoria escolar. Además, en la provincia de Los Ríos operan transferencias monetarias condicionadas y becas locales de alimentación escolar que atenúan la restricción presupuestaria de los hogares para gastos educativos. Por ello, el efecto de la pobreza tiende a canalizarse a través de otros mecanismos ya capturados en el modelo —por ejemplo, la necesidad de que el menor trabaje— o bien queda parcialmente neutralizado por estas políticas de compensación. Así, la “no

significancia” observada no implica irrelevancia de la pobreza, sino que su influencia se expresa de forma indirecta o está amortiguada por estas intervenciones sociales.

La variable que sí presentó un efecto notable y negativo fue el indicador de trabajo reciente, consistente con la literatura que describe al trabajo infantil como una actividad que compite directamente con la educación. Sin embargo, su influencia no eclipsó la importancia de la edad como predictor principal. Esto refuerza la idea de que, si bien el trabajo puede ser una manifestación de vulnerabilidad, no es el único mecanismo que lleva a los adolescentes a dejar la escuela.

En cuanto al sexo y la zona de residencia, los resultados no mostraron diferencias estadísticamente significativas, lo cual puede deberse a mejoras en la cobertura educativa rural o a dinámicas de género más equilibradas en la provincia. Aun así, se recomienda explorar con mayor profundidad si existen brechas ocultas en otras dimensiones (como la calidad educativa, la permanencia o el rendimiento) no captadas por este análisis binario. En términos de política pública, los hallazgos apuntan a la necesidad de intervenciones focalizadas por tramo etario, particularmente entre los 12 y 15 años. Se sugiere fortalecer programas de acompañamiento escolar, prevención del abandono en secundaria, y sistemas de alerta temprana que combinen información de asistencia, edad y antecedentes laborales. Además, dado el bajo peso estadístico de las variables económicas, las estrategias deberían mirar más allá del ingreso y considerar factores psicosociales, motivacionales y pedagógicos.

Finalmente, esta investigación pone en evidencia la utilidad de los métodos estadísticos aplicados sobre microdatos oficiales para responder preguntas locales relevantes. La provincia de Los Ríos carece de estudios cuantitativos de este tipo, lo cual hace que este análisis no solo contribuya a la literatura nacional, sino que sirva como insumo para los gobiernos locales, el sistema educativo y los actores comunitarios interesados en revertir el abandono escolar con evidencia.

REFERENCIAS BIBLIOGRÁFICAS

- Alhassan, A. H. I., & Hassan, Z. B. (2024). Understanding the Factors Behind Dropout Decisions: A Study of Primary School Students in Red Sea State, Sudan. *International Journal of Research and Innovation in Social Science*, VIII(IIIS), 5023-5034. <https://doi.org/10.47772/ijriss.2024.803373s>
- Binnis, M., & Davalos, A. (2023). Does Internet Access Mitigate the Impact of COVID-19 on School Re-Enrollment Rates? Evidence from Mexico. *Creative Education*, 14(11), 2131-2151. <https://doi.org/10.4236/ce.2023.1411136>
- Bracco, J., Ciaschi, M., Gasparini, L., Marchionni, M., & Neidhöfer, G. (2025). The Impact of COVID-19 on Education in Latin America: Long-Run Implications for Poverty and Inequality. *Review of Income and Wealth*, 71(1). <https://doi.org/10.1111/roiw.12687>
- De Vasconcelos, A. N., Freires, L. A., Loureto, G. D. L., Fortes, G., Da Costa, J. C. A., Torres, L. F. F., Bittencourt, I. I., Cordeiro, T. D., & Isotani, S. (2023). Advancing school dropout early warning systems: The IAFREE relational model for identifying at-risk students. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1189283>
- Dey, D., Haque, Md. S., Islam, Md. M., Aishi, U. I., Shammy, S. S., Mayen, Md. S. A., Noor, S. T. A., & Uddin, Md. J. (2025). The proper application of logistic regression model in complex survey data: A systematic review. *BMC Medical Research Methodology*, 25(1). <https://doi.org/10.1186/s12874-024-02454-5>
- Díaz-Coto, S., Martínez-Camblor, P., & Corral-Blanco, N. (2024). Unified ROC Curve Estimator for Diagnosis and Prognosis Studies: The sMSROC Package. *The R Journal*, 15(4), 129-149. <https://doi.org/10.32614/rj-2023-087>

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R.

G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C.,

Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., &

Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27-46.

<https://doi.org/10.1111/j.1600-0587.2012.07348.x>

Howell-Moroney, M. (2024). Inconvenient truths about logistic regression and the

remedy of marginal effects. *Public Administration Review*, 84(6), 1218-1236.

<https://doi.org/10.1111/puar.13786>

Instituto Nacional de Estadística y Censos. (2024). *Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU)* [Dataset].

<https://www.ecuadorencifras.gob.ec/enemdu-2024/>

Inter-American Development Bank. (2024). *The State of Education in Latin America and the Caribbean 2024*. Inter-American Development Bank.

<https://publications.iadb.org/publications/english/document/The-state-of-education-in-Latin-America-and-the-Caribbean-2024.pdf>

International Labour Organization. (2023). *Issue paper on child labour and education exclusion among indigenous children*. International Labour Organization.

[https://www.ilo.org/sites/default/files/2024-](https://www.ilo.org/sites/default/files/2024-04/Child_Labour_and_Indigenous_children_2023.pdf)

[04/Child_Labour_and_Indigenous_children_2023.pdf](https://www.ilo.org/sites/default/files/2024-04/Child_Labour_and_Indigenous_children_2023.pdf)

Iparragirre, A., Barrio, I., & Arostegui, I. (2023). Estimation of the ROC curve and the area under it with complex survey data. *Stat*, 12(1).

<https://doi.org/10.1002/sta4.635>

Makowski, D., Ben-Shachar, M. S., Wiernik, B. M., Patil, I., Thériault, R., & Lüdecke, D. (2025). *modelbased*: An R package to make the most out of your statistical

models through marginal means, marginal effects, and model predictions.

Journal of Open Source Software, 10(109), 7969.

<https://doi.org/10.21105/joss.07969>

Nigri, A., Bilancia, M., Cafarelli, B., & Magro, S. (2025). *Modelling higher education dropouts using sparse and interpretable post-clustering logistic regression*

(Versión 1). arXiv. <https://doi.org/10.48550/ARXIV.2505.07582>

Niu, L. (2020). A review of the application of logistic regression in educational research: Common issues, implications, and suggestions. *Educational Review*, 72(1), 41-67. <https://doi.org/10.1080/00131911.2018.1483892>

Qiu, Q., Li, S., Chen, Y., Yan, X., Yang, S., Qiu, S., Peng, A., & Chen, Y. (2023). Development, assessment and validation of a novel prediction nomogram model for risk identification of tracheobronchial tuberculosis in patients with pulmonary tuberculosis. *BMJ Open Respiratory Research*, 10(1), e001781. <https://doi.org/10.1136/bmjresp-2023-001781>

Reibnegger, G., & Schrabmair, W. (2014). Optimum binary cut-off threshold of a diagnostic test: Comparison of different methods using Monte Carlo technique. *BMC Medical Informatics and Decision Making*, 14(1). <https://doi.org/10.1186/s12911-014-0099-1>

Sarcos González, J. A. (2025). Identificación y Caracterización de Perfiles de Inactividad Laboral en Ecuador mediante Análisis de Componentes Principales y Clústeres K-means. *Ciencia Latina Revista Científica Multidisciplinar*, 8(6), 7568-7586. https://doi.org/10.37811/cl_rcm.v8i6.15451