

ESTIMACIÓN DE LA PROBABILIDAD DE POBREZA POR INGRESOS EN ECUADOR MEDIANTE UN MODELO LOGIT, 2022

ESTIMATION OF THE PROBABILITY OF INCOME POVERTY IN ECUADOR USING A LOGIT MODEL, 2022

Paúl Marín-Vélez¹, Livino Armijos-Toro²

{paulismael18@hotmail.com, livinoa@gmail.com}

Fecha de recepción: 25/02/2026 / Fecha de aceptación: 28/03/2026 / Fecha de publicación: 31/03/2026

RESUMEN: La pobreza es un fenómeno social que ha venido afectando a la población ecuatoriana en el transcurso del tiempo. En este sentido, la presente investigación se enfoca en estimar la probabilidad que tienen los individuos de llegar a ser pobres en el Ecuador en el año 2022, usando un modelo probabilístico logístico (logit). Mencionado modelo tiene como variable dependiente la Pobreza por Ingresos y como variables independientes: Jefe de Hogar, Zona Rural, Ingreso Per Cápita y Edad. Para los cálculos matemáticos y estadísticos, se emplea la base de datos del Instituto Nacional de Estadísticas y Censos (INEC), llamada Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) Acumulada del año 2022. Utilizando el modelo logit, se realiza un análisis muestral donde sus parámetros estimados son coherentes con la teoría económica-social; también, se calculan los efectos marginales de las variables predictoras mostrando que son lógicos y estadísticamente significativos, además, se evalúa la razón de probabilidad de un evento o los OddsRatios exponiendo que las variables independientes generan efectos en la probabilidad del suceso, finalmente se demuestra que el modelo logit muestral tiene una gran capacidad predictiva mediante la curva ROC. Asimismo, se realiza un análisis poblacional agregando al modelo muestral el factor de expiación (valor construido por el INEC), generando un modelo logit poblacional, con el propósito de comparar los resultados muestrales con los poblacionales y observar que tan efectivo es el logit muestral, pretendiendo tener estimaciones semejantes en los dos modelos. El nuevo modelo poblacional indica que dos de las cuatro variables explicativas del estudio muestral son estadísticamente significativas, dichas variables son: Ingreso Per Cápita y Edad. Consiguientemente, las dos variables que conforman el modelo logit poblacional enseñan efectos marginales razonables, sus OddsRatios indican efectos sobre la pobreza, y lo más importante, sus resultados son muy semejantes al logit muestral. Es así, que los modelos usados en el estudio, explican la probabilidad que tienen los individuos de ser pobres por ingresos en el Ecuador año 2022;

¹Universidad Politécnica Estatal del Carchi, Posgrado, Av. Universitaria y Antisana, Tulcán, 040101; <https://orcid.org/0009-0002-0975-9200>, 0992747085.

²Universidad Internacional del Ecuador, Business School, Av. Simón Bolívar y Av. Jorge Fernández, Quito, 170411; <https://orcid.org/0000-0001-8553-536X>, 0992669890.

además, se identifican las variables que conducen a la pobreza por rentas y poseen gran capacidad predictiva.

Palabras Clave: Pobreza, probabilidad, modelo probabilístico logístico, población, muestra, parámetros, efectos marginales, OddsRatios, Curva ROC

ABSTRACT: Poverty is a social phenomenon that has affected the Ecuadorian population over time. This research focuses on estimating the probability of individuals becoming poor in Ecuador in 2022, using a logistic probability model (logit). The dependent variable of this model is Income Poverty, and the independent variables are Head of Household, Rural Area, Per Capita Income, and Age. For the mathematical and statistical calculations, the National Institute of Statistics and Censuses (INEC) database, specifically the 2022 National Survey of Employment, Unemployment, and Underemployment (ENEMDU), is used. Using the logit model, a sample analysis is performed, and the estimated parameters are consistent with socio-economic theory. The marginal effects of the predictor variables are also calculated, showing that they are logical and statistically significant. Furthermore, the odds ratios of an event are evaluated, demonstrating that the independent variables influence the probability of the event. Finally, the ROC curve demonstrates the high predictive capacity of the sample logit model. Additionally, a population analysis is performed by adding the atonement factor (a value constructed by the INEC) to the sample model, generating a population logit model. This aims to compare the sample results with the population results and observe the effectiveness of the sample logit model, with the goal of obtaining similar estimates in both models. The new population model indicates that two of the four explanatory variables in the sample study are statistically significant: Per Capita Income and Age. Consequently, the two variables that make up the population logit model show reasonable marginal effects, their odds ratios indicate effects on poverty, and most importantly, their results are very similar to the sample logit model. Thus, the models used in the study explain the probability that individuals have of being income poor in Ecuador in 2022; in addition, the variables that lead to income poverty are identified and have great predictive capacity.

Keywords: Poverty, probability, probabilistic logistic model, population, sample, parameters, marginal effects, Odds Ratios, ROC Curve

INTRODUCCIÓN

Entender la pobreza ha sido un tema de debate a lo largo del tiempo en la sociedad. Los antiguos filósofos planteaban que el Estado debe corregir la injusticia que genera la pobreza con el objetivo de disminuirla. La perspectiva económica clásica (1776) proponía que una sociedad es más rica que otra si posee mayor mano de obra, a más individuos, mayor producción y menos pobreza. En 1800 se complementa la idea clásica con el argumento, a mayor riqueza, mayor satisfacción de las necesidades. El pensamiento económico neoclásico, exponía el crecimiento de la riqueza como un proceso paulatino de pequeños cambios incrementales, creando la teoría de la utilidad. A principios del siglo XIX se mira a la pobreza como la carencia de recursos materiales, comparando

los ingresos con los gastos de una entidad. En los años treinta del siglo pasado aparece un concepto ligado a la pobreza que es la desigualdad de la riqueza. A finales del siglo XIX e inicio del siglo XX se mira a la pobreza no solo como un problema económico sino social, se comienza a ver a la misma de forma multidimensional indicando que es una condición que perjudica el desarrollo mental, físico, emocional y espiritual de los individuos (1).

En la actualidad, la pobreza afecta a gran parte de los habitantes del mundo, una de cada tres personas en el planeta vive en condiciones de pobreza, hasta los países más ricos padecen este mal. A pesar de los esfuerzos que venían haciendo los Estados nacionales por disminuir los porcentajes de pobreza, estos se agudizaron con la pandemia de la COVID-19 en los años 2020-2021, provocando que más de 700 millones de personas vivan en extrema pobreza en el planeta (2). De acuerdo a la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO), la pobreza es un problema de carácter estructural de todas las sociedades sin excepción, ubicándose dentro de los objetivos de desarrollo sostenible, requiriendo eliminarla en y entre países hasta el año 2030 (3). Asimismo, el Ecuador no se excluye del inconveniente de la pobreza, según al Instituto Nacional de Estadísticas y Censos (INEC), en el primer semestre del año 2022, las personas pobres por ingresos del país representaron un 25% de la población, cifra alta tomando en cuenta que estamos en un proceso de recuperación económica por la reactivación de la economía luego de la pandemia. La pobreza extrema por ingresos representó el 10.7% nacional (4).

En este sentido, observando la importancia del problema de la pobreza en las sociedades, esta investigación plantea como pregunta central: ¿Cuál es la probabilidad que tienen las personas en el Ecuador de llegar a la pobreza por ingresos en el año 2022?; y preguntas específicas: ¿Qué factores tienen mayor posibilidad de conducir a los individuos a la pobreza por ingresos?, ¿Se pueden predecir los niveles de pobreza por ingresos? Consecuentemente, la variable de estudio es la Pobreza por Ingresos que se determina actualizando la línea oficial de la pobreza en función del consumo, basados en el índice de Precios al Consumidor (IPC), así, una persona es pobre por ingresos si su renta total per cápita es menor a la línea de pobreza (5).

Consiguientemente, el artículo usa un modelo logit para alcanzar el objetivo principal que es: Estimar la probabilidad que tienen las personas de ser pobres por ingresos en el Ecuador en el año 2022, y llegar a sus objetivos específicos que son: Encontrar los factores que tienen mayor probabilidad de conducir a las personas a la pobreza por ingresos en el país y poder predecirla.

MATERIALES Y MÉTODOS

Materiales

El artículo emplea la base de datos del INEC, denominada Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) Acumulada del año 2022, la misma que cuenta con 139 variables de las cuales se usan 9 (6).

El INEC, en su Boletín Técnico N° 05-2023 ENEMDU muestra que la base de datos arriba mencionada tiene un tamaño muestral anual de 108192 viviendas, y 358096 personas (7).

El Diseño Muestral del ENEMDU posee un dominio de estudio y repetitividad nacional (urbana y rural); las ciudades de Quito, Guayaquil, Cuenca, Machala, Ambato y 24 provincias del país. El universo de investigación son las personas de 5 años y más de edad que viven en las residencias del país, excepto los individuos que habitan en casas colectivas, indigentes etc. La unidad de observación son todas las viviendas ocupadas que están en el territorio ecuatoriano. La unidad de análisis con respecto a los indicadores laborales son los individuos de 15 años y más. Finalmente, el tipo de muestreo es probabilístico estratificado bietápico (8).

VARIABLES UTILIZADAS EN LA INVESTIGACIÓN

La variable predicha y las variables predictoras usadas son las siguientes:

Variable dependiente

Pobreza por Ingresos (Y_i); Según el INEC, una persona es considerada pobre por ingresos si tiene una renta per cápita menor a \$88,72 que representa la línea de la pobreza (5). En consecuencia, esta variable toma el valor de 1 si la persona es "POBRE" y 0 si la persona es "NO POBRE", siendo la misma una variable cualitativa dicotoma.

$$Y_i = \begin{cases} 1 = \text{"POBRE"} \\ 0 = \text{"NO POBRE"} \end{cases}$$

VARIABLES INDEPENDIENTES

Las variables de control se basan en las características socioeconómicas de los individuos que residen en el Ecuador.

Narváez, Parra y Burgos; consideran algunas variables que ayudan a entender la pobreza, se toman las siguientes (9):

Jefe de Hogar (X_1); Ocupa el valor de 1 si la persona se identifica como "Jefe", y 0 caso contrario, siendo la misma una variable cualitativa categórica.

$$X_1 = \begin{cases} 1 = \text{"Jefe"} \\ 0 = \text{"Conyugue"} \\ 0 = \text{"Hijo o Hija"} \\ 0 = \text{"Yerno o nuera"} \\ 0 = \text{"Nieto o nieta"} \\ 0 = \text{"Padres o suegros"} \\ 0 = \text{"Otros parientes"} \\ 0 = \text{"Empleado(a) doméstico(a)"} \\ 0 = \text{"Otros no parientes"} \end{cases}$$

Sexo Femenino (X_2); Acepta el valor de 1 si el individuo encuestado se considera "Mujer" y 0 si se califica "Hombre". Variable cualitativa dicotómica.

$$X_2 = \begin{cases} 0 = \text{"Hombre"} \\ 1 = \text{"Mujer"} \end{cases}$$

Edad (X_3); Edad en años. Variable cuantitativa.

Del mismo modo Pucutay, plantea varias variables para analizar la pobreza, y se escogieron estas (10):

Estado Civil, Casado (a) (X_4); Persona casado (a), recibe el valor de 1 si cumple esta condición, y 0 si no cumple la misma. En este sentido es una variable categórica.

$$X_4 = \begin{cases} 1 = \text{"Casado(a)"} \\ 0 = \text{"Separado(a)"} \\ 0 = \text{"Divorciado(a)"} \\ 0 = \text{"Viudo(a)"} \\ 0 = \text{"Unión libre"} \\ 0 = \text{"Soltero(a)"} \end{cases}$$

Ingreso Per Cápita (X_5); Ingreso de las personas miembros de un hogar, variable cuantitativa.

Algunos autores proponen diversas variables para el estudio de la pobreza, se emplean las que se presentan a continuación (11):

Nivel de Educación Alcanzado, Secundaria (X_6); "Secundaria" como nivel de educación alcanzado recibiendo el valor de 1, y 0 si no lo es. Variable cualitativa categórica.

$$X_6 = \begin{cases} 0 = \text{"Ninguno"} \\ 0 = \text{"Centro de alfabetización"} \\ 0 = \text{"Jardín de infantes"} \\ 0 = \text{"Primaria"} \\ 0 = \text{"Educación Básica"} \\ 1 = \text{"Secundaria"} \\ 0 = \text{"Educación Media"} \\ 0 = \text{"Superior no universitario"} \\ 0 = \text{"Superior Universitario"} \\ 0 = \text{"Post - grado"} \end{cases}$$

Área Rural (X_7); Persona que vive en el área rural, tomando el valor de 1, y cero si vive en el sector urbano. Variable cualitativa dicotómica.

$$X_7 = \begin{cases} 0 = \text{"Urbana"} \\ 1 = \text{"Rural"} \end{cases}$$

Factor de Expansión (X_8); Variable cuantitativa que permite proyectar los resultados al total de la población.

Métodos

El método que se utiliza en la investigación para medir la probabilidad de que los individuos sean pobres en el Ecuador en el año 2022, es el modelo de probabilidad logit.

Modelos probabilísticos

Gujarati y Porter, plantean que en los modelos probabilísticos la variable respuesta (Y_i) es cualitativa, es decir, la incógnita regresada es policótoma (categoría múltiple). En consecuencia, el objetivo de los modelos de variable explicada categórica es hallar la probabilidad de que un evento ocurra. En este sentido, existen cuatro modelos de probabilidad que se usan para una variable de respuesta binaria que son (12):

1. Modelo Lineal de Probabilidad (MLP).
2. Modelo Tobit.
3. Modelo Logit.
4. Modelo Probit.

El estudio no utiliza el MLP, y el modelo tobit por sus características teóricas funcionales, y el modelo probit, debido al extenso análisis que se debería realizar. Sin embargo, se hace un estudio teórico de mencionados métodos y porqué de su no uso.

Modelo lineal de probabilidad

No normalidad de las perturbaciones (u_i): al ser un modelo de repuesta binaria (Y_i), las perturbaciones (u_i) siguen una distribución de probabilidad Bernoulli, es decir, los errores no siguen una distribución normal dificultando la estimación de los parámetros mediante el método de mínimos cuadrados ordinarios (MCO). Sin embargo, si las observaciones (n) tienden al infinito, los estimadores puntuales de MCO seguirán una distribución normal y serán insesgados.

Varianzas heteroscedásticas de las perturbaciones: en la distribución Bernoulli, la media (p) es la probabilidad de éxito de un evento y la varianza es igual a $p(1 - p)$. Consecuentemente, la varianza es una función de la media, generando que las perturbaciones tengan una varianza heteroscedástica.

Se sabe que si existe heteroscedasticidad los parámetros estimados por MCO son ineficientes porque no tienen varianza mínima, pero el problema es superable si se divide al modelo por un ponderado $\sqrt{w_i}$ y luego se calculan los parámetros por mencionado método.

No cumplimiento de $0 \leq E(Y_i/X_i) \leq 1$: MCO no acepta una desigualdad como restricción generando un problema en la estimación de los parámetros.

Se cuestiona al R^2 como medida de bondad de ajuste: el R^2 mide que tan cerca están las observaciones de la línea de regresión, pero en el MLP dado un valor de X, Y se encuentra entre 0 y 1, es decir que cualquier observación se va a encontrar o bien en el eje de las X o Y (gráfica 1), indicando un ajuste limitado y un R^2 bajo.

Debido a los problemas de estimación mencionados, se requiere una función no lineal que muestre que a medida que la variable independiente X_i se incrementa, la probabilidad condicional de ocurrencia de un evento $P_i = E(Y = 1/X)$ también aumenta, y a medida que X_i

disminuya, la probabilidad también se reduzca sin exceder en ningún momento el intervalo entre 0 y 1.

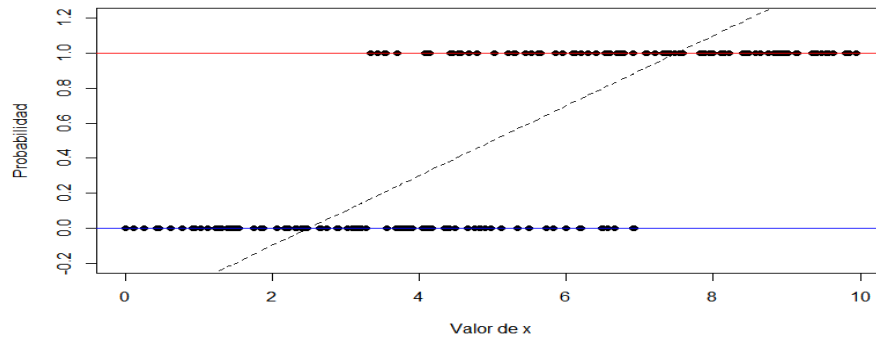


Figura 1. Modelo Lineal de Probabilidad.

Modelo Tobit

El método tobit trabaja con dos tamaños muestrales (n_1 y n_2), donde la primera muestra tiene información de las variables independientes y la variable dependiente, la segunda muestra tiene información de las variables exógenas pero no de la variable endógena. Consecuentemente, el estudio planteado, cuenta con una sola muestra y con información completa de todas las variables.

Modelo Logit

Guevara y Uribe, plantean que el logit es un modelo no lineal basado en una función logística. Se emplea para calcular la probabilidad condicional de un evento de una variable categórica (binomial, multinomial) en función de variables cualitativas o cuantitativas. Este modelo establece que las probabilidades calculadas se encuentran en el intervalo de [0 y 1] y la variable dependiente toma valores de 0 = *desacierto* y 1 = *acierto* (13).

Partiendo de la función de distribución acumulada (FDA) logística, tenemos que la probabilidad de ocurrencia de un acontecimiento está dada por la siguiente igualdad:

$$P_i = \frac{1}{1+e^{-z_i}} = \frac{e^{z_i}}{1+e^{z_i}} \quad (1)$$

Empleando la función lineal del MLP se conoce el valor de z_i

$$z_i = \alpha + \beta_2 X_{i2} + \dots + \beta_k X_{ik} \quad (2)$$

Reemplazando el valor de z_i en la función logística se tiene

$$P_i = \frac{e^{(\alpha + \beta_2 X_{i2} + \dots + \beta_k X_{ik})}}{1 + e^{(\alpha + \beta_2 X_{i2} + \dots + \beta_k X_{ik})}} \quad (3)$$

Se observa que la función no es lineal en variables y parámetros, por lo que no se pueden estimar los parámetros por el método de MCO y se emplea el procedimiento de máxima verosimilitud cuando se trabaja con datos individuales (el caso de estudio).

Efectos Marginales (EM) en el modelo logit

Para mencionados autores la no linealidad en parámetros provoca que los mismos no sean interpretados directamente después de su cálculo, así que para llegar a su valor real se deriva parcialmente la función, este resultado se conoce como los efectos marginales (EM).

Iniciando con la probabilidad de éxito de un evento:

$$P_i = \frac{e^{z_i}}{1 + e^{z_i}}$$

Derivando

$$EM\beta_k = \frac{(e^{z_i} * \beta_k) * (1 + e^{z_i}) + (e^{z_i}) * (e^{z_i} * \beta_k)}{(1 + e^{z_i})^2} \quad (4)$$

Queda

$$EM\beta_k = \frac{e^{z_i}}{(1 + e^{z_i})^2} * \beta_k \quad (5)$$

Usando las probabilidades de éxito y fracaso la igualdad (5) resulta

$$(1 - P_i) = 1 - \frac{e^{z_i}}{1 + e^{z_i}} = \frac{1}{(1 + e^{z_i})} \text{ probabilidad de fracaso (6)}$$

$$EM\beta_k = \frac{e^{z_i}}{(1 + e^{z_i})} * \frac{1}{(1 + e^{z_i})} * \beta_k \quad (7)$$

$$EM\beta_k = P_i * (1 - P_i) * \beta_k \quad (8)$$

A partir de aquí, los parámetros se interpretan de la forma normal.

Efectos Marginales Globales en el modelo logit

Miden los efectos marginales en forma global. Para las variables cuantitativas se obtiene el promedio y para las incógnitas cualitativas se adquiere la proporción para cada categoría.

Modelo Probit

Para Guevara y Uribe, el probit (también llamado normit) es otro modelo probabilístico no lineal usado para medir posibilidad condicional de que un suceso acontezca de una variable categórica binaria. Se basa en una función de distribución acumulada (FDA) normal. Al igual que el logit, en

el normit las probabilidades calculadas se encuentran en el intervalo de [0 y 1]. En este sentido, para observar los efectos marginales de los parámetros se usa la función de densidad de probabilidad (fdp), y para predecir se usa la FDA (13).

La función de distribución normal transformada a un modelo equivalente de la función lineal es el índice de convergencia (I_i) o variable latente:

$$I_i = \alpha + \beta_2 X_{i2} + \dots + \beta_k X_{ik} \quad (9)$$

Se plantea un nivel crítico (I_i^*) menor al índice de convergencia para encontrar la probabilidad de éxito de un evento. Es decir, si el índice de convergencia es mayor al umbral establecido muestra el éxito del evento:

$$P(I_i^* \leq I_i) \quad (10)$$

En consecuencia, se dice que:

$$\begin{aligned} P_i &= P(Y = 1 / X) = P(I_i^* \leq I_i) \\ &= P(Z_i \leq \alpha + \beta_2 X_{i2} + \dots + \beta_k X_{ik}) \\ &= F(\alpha + \beta_2 X_{i2} + \dots + \beta_k X_{ik}) \quad (11) \end{aligned}$$

Reemplazando en una distribución de probabilidad normal estándar queda:

$$F(I) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{I_i} e^{-\frac{z^2}{2\sigma^2}} \cdot dZ \quad (12)$$

Y sustituyendo mencionado índice por si igualdad:

$$F(I) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\alpha + \beta_2 X_{i2} + \dots + \beta_k X_{ik}} e^{-\frac{z^2}{2\sigma^2}} \cdot dZ \quad (13)$$

La identidad (13) representa la FDA, donde la posibilidad de ocurrencia de un evento se mide por el área bajo la curva normal estándar.

Para obtener los valores unitarios de la probabilidad de que un evento suceda, se usa la función inversa del índice de convergencia:

$$I^{-1} = \alpha + \beta_2 X_{i2} + \dots + \beta_k X_{ik} \quad (14)$$

Esta función sirve para calcular los parámetros del modelo y sus efectos marginales.

Efectos Marginales (EM) en el modelo probit

Al igual que en el logit, los parámetros del modelo probit no son lineales porque son calculados por el método de máxima verosimilitud y su interpretación no es directa, en consecuencia se deben linealizar para su análisis.

Los efectos marginales en el modelo probit muestran la probabilidad de ocurrencia en ese punto, convirtiéndose en una función de densidad de probabilidad (fdp).

$$EM\hat{\beta}_k = \frac{\partial(P_i)}{\partial(X_i)} = f(\hat{\alpha} + \hat{\beta}_2X_{i2} + \dots + \hat{\beta}_kX_{ik}) * \hat{\beta}_k \quad (15)$$

Efectos Marginales Globales en el modelo probit

En el modelo probit también se puede obtener una medida global de los efectos marginales.

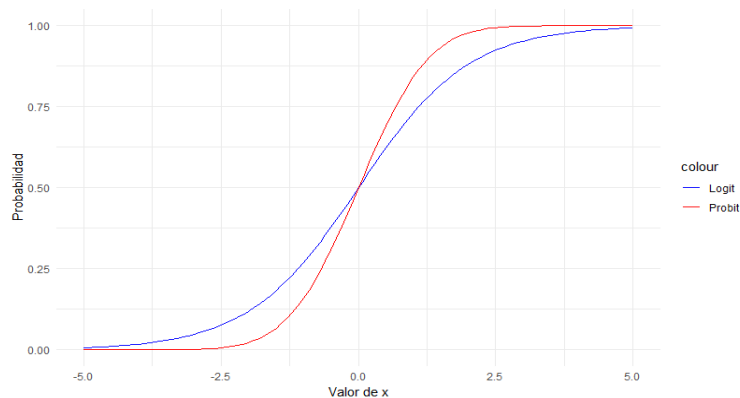


Figura 2. Modelo Logit y Probit.

Marco experimental

En el modelo probabilístico logit, la variable dependiente es la Pobreza por Ingresos, y las variables independientes son: Jefe de Hogar, Sexo Femenino, Edad, Estado Civil (Casado/a), Ingreso per Cápita, Nivel de Educación Alcanzado (Secundaria), y Área Rural para el análisis muestral.

Para el análisis nacional se usan las mismas variables y el Factor de Expansión creado por el INEC.

La base de datos empleada del INEC llamada ENEMDU, como se mencionó anteriormente cuenta con una un tamaño muestral anual de 108192 viviendas y 358096 personas, donde el sujeto de estudio es el individuo.

Un paso previo al uso de la información contenida en la base referida arriba, fue el tratamiento computacional de datos, donde se encontraron las siguientes novedades:

Datos faltantes (NA): Para Segura y Torres (14), los NA son valores ausentes en las bases de datos, generando errores en el análisis estadístico. Las causas de su inexistencia son varias, las más frecuentes pueden ser que los encuestados no responden algunas preguntas, errores en los instrumentos de medición, etc. Para la solución de este inconveniente dichos autores recomiendan la eliminación de los NA o imputación.

En la investigación se eligió imputar datos o completar los mismos, porque las variables Ingreso per Capita y Pobreza contaban con 2326 NA cada una en diferentes filas, y al eliminarlos se excluían demasiada información de otras variables.

El método de imputación usado se conoce como los k vecinos más cercanos (kNN o k-nearest neighbors). Se basa en que las características de una observación x son afines a los de su vecindad. En este sentido se propone sustituir al dato NA por el vecino más cercano en función de una distancia elegida (15).

La razón principal de uso de la técnica kNN es porque predice datos cualitativos y cuantitativos (16).

Datos Atípicos (outliers): Son observaciones con rasgos diferentes a las demás (valores grandes o pequeños), pueden generar resultados no confiables. Para usar los valores atípicos en un estudio se debe analizar el contexto del fenómeno y validar esa información (17). Son varias las causas que provocan los outliers, las más comunes son: errores en el ingreso de información, ocurrencia de un fenómeno fuera de lo común en la variable de estudio, etc.

La técnica que se usó para detectar datos atípicos es el diagrama de caja, basado en la obtención de la mediana, los cuartiles 1 y 3 y el rango intercuartílico. Es así que; se dibuja una caja que contiene el 50% de las observaciones centrales, luego se detecta la mediana y se traza una recta vertical, los extremos de la caja son el primer y tercer cuartiles. Seguidamente se calculan los límites usando el rango intercuartílico donde los datos que se están fuera de estas fronteras son considerados atípicos. Luego se trazan unas líneas punteadas (bigotes) desde los extremos de la caja hasta los límites. Finalmente con un punto se señalan los datos atípicos (18).

De acuerdo a lo expuesto arriba se observa (ver Figura 3), que la variable ingreso per cápita tiene valores atípicos, sin embargo, se van a utilizar estos porque los datos presentados por el INEC son oficiales para el Ecuador y son observaciones que se han repetido a nivel del año.

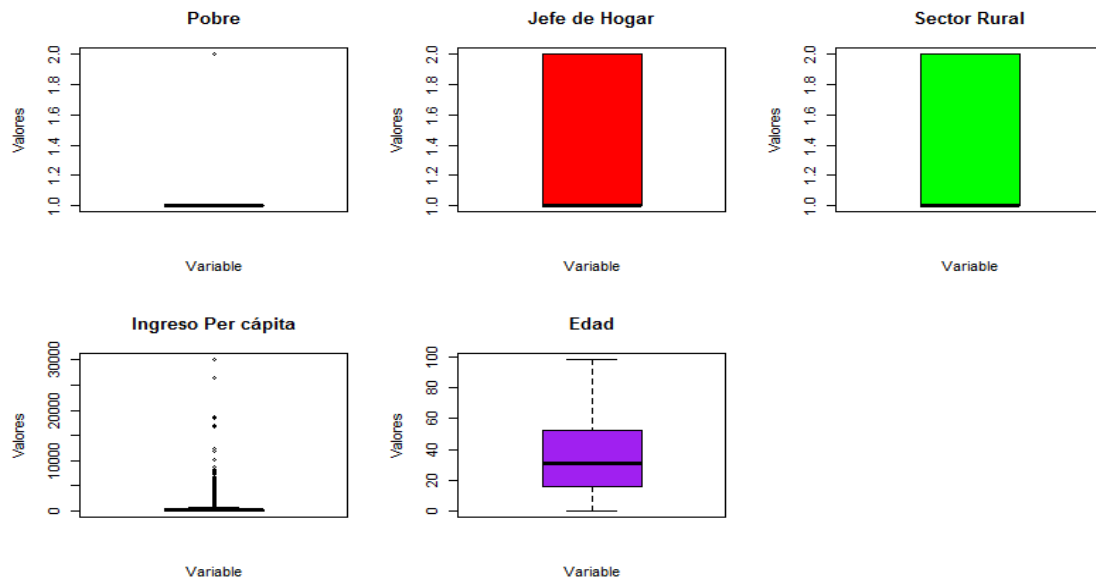


Figura 3. Valores atípicos de las variables.

Luego de la sustitución de los NA, y el análisis de los outliers de la base de datos, se divide la misma de la siguiente forma: 80% (286476) de las observaciones se usan para estimación y el 20% (71620) para predicción, con la intención de evidenciar que tan bien predice el modelo con otras observaciones.

La elección de los datos para estimación y predicción se realizó de forma aleatoria, usando números aleatorios generados en el programa R y RStudio. Entendiendo por números aleatorios a una sucesión de números donde no se conoce cuál será el siguiente de ellos y si a dicha serie numérica la sometemos a cualquier prueba de aleatoriedad la supera sin ningún problema (19).

RESULTADOS

En esta sección se presentan los principales resultados de la investigación, se utiliza el programa computacional R y RStudio para todos los cálculos matemáticos y estadísticos.

Las pruebas de hipótesis con sus diferentes fines de validación tienen un nivel de significancia del 5%, dado que es el valor más usado en la mayoría de los estudios estadísticos.

Modelo Logit muestral

Parámetros

Como se observa en la tabla 1, los parámetros de las variables “Jefe de Hogar”, “Zona Rural”, “Ingreso per Cápita” y “Edad”, son estadísticamente significativos para decir que el valor de los mismos son diferentes de cero, ya que el valor p es menor que el nivel de significancia del 5%.

Efectos Marginales

Los efectos marginales de las variables son:

Jefe de Hogar: Por un incremento de una persona como Jefe de Hogar, la probabilidad de ser pobre por ingresos en el Ecuador disminuye en un 0,2%, manteniendo las demás variables constantes.

Zona Rural: Por un aumento de una persona en la Zona Rural, la probabilidad de ser pobre de aquel individuo en el país se incrementa en un 0.52%, con las demás variables constantes.

Ingreso Per Cápita: Por un incremento de un dólar en los ingresos de las personas que viven en el Ecuador, la probabilidad de ser pobre se reduce en un 0,34%, manteniendo las demás variables constantes.

Edad: Por un aumento de un año en la edad de los individuos que viven en el país, la probabilidad de ser pobre baja en un 0.02%, con las demás variables constantes (ver Tabla 1).

OddsRatios

Se evidencian los siguientes resultados:

Jefe de Hogar: Dado un incremento de una persona como Jefe de Hogar, la probabilidad de ser pobre por ingresos se multiplica aproximadamente por 0.899. Entendiendo que la variable Jefe de Hogar está relacionada con una disminución de la posibilidad de ocurrencia del evento.

Zona Rural: Por un aumento de una persona en la Zona Rural, la posibilidad de llegar a la pobreza se multiplica por 1.316. Es decir que la variable Zona Rural está relacionada con un incremento en la probabilidad de ocurrencia del suceso.

Ingreso Per Cápita: Dado un incremento de un dólar en los ingresos de las personas que viven en el Ecuador, la probabilidad de ser pobre se multiplica 0.835. En consecuencia, los ingresos per cápita están relacionados con una disminución de la posibilidad de ocurrencia del evento.

Edad: Por un aumento de un año en la edad de los individuos que viven en el país, la posibilidad de ser pobre por renta se multiplica por 0.989. Consecuentemente, la Edad relacionada con una disminución en la probabilidad de ocurrencia del suceso (ver tabla 1).

Evaluación del Modelo

Significancia Individual: El nivel de significancia individual se analiza mediante la prueba z, donde todos los parámetros del modelo son estadísticamente significativos (ver Tabla 1).

Significancia Global: El modelo en conjunto es estadísticamente significativo, porque el nivel significancia global del modelo es igual a 225544,43, con un valor $p = 0.000$ (ver Tabla 1).

Bondad de Ajuste

El $R^2 McFadden$, es de 0.87 (ver Tabla 1).

El Log Likelihood (MV) o logaritmo de la función de verosimilitud = -17738.47 (ver Tabla 1).

Multicolinealidad

En el modelo logit de estudio, todas las variables tienen un FIV menor de 1.5, lo que indica que no tiene problemas de multicolinealidad (ver Tabla 1).

Finalmente la probabilidad promedio de ser pobre en la muestra es de 19.30%.

Tabla 1. Resultados del modelo logit muestral.

Estimación				
$Pobreza\ por\ Ingresos_i = 15.733 - 0.107(Jefe\ de\ Hogar_i) + 0.275(Zona\ Rural_i) - 0.0034(Ingreso\ Per\ Cápita_i) - 0.011(Edad_i)$				
	Estimate	Std. Error	z value	Pr(> z)
Intercepto	15.733	0.121	130.052	0.0000
Jefe de Hogar	-0.107	0.038	-2.833	0.0046
Zona Rural	0.275	0.028	9.738	0.0000
Ingreso Per Cápita	-0.180	0.001	-133.012	0.0000
Edad	-0.011	0.001	-14.274	0.0000
Efectos Marginales				
	Efectos Marginales	%		
Jefe de Hogar	-0.002	-0.2		
Zona Rural	0.0052	0.52		

**ESTIMACIÓN DE LA PROBABILIDAD DE POBREZA POR INGRESOS EN ECUADOR MEDIANTE UN MODELO LOGIT,
2022**

Ingreso Per Cápita	-0.0034	-0.34
Edad	-0.0002	-0.02

OddsRatios

	OddsRatios	%
Jefe de Hogar	0.899	-10.122
Zona Rural	1.316	31.625
Ingreso Per Cápita	0.835	-16.513
Edad	0.989	-1.051

Evaluación del Modelo

Significancia Global	$\chi^2(4)$	245544.34	p = 0.00
Bondad de Ajuste			
	Pseudo-R ² (McFadden)	0.87	
	Log Likelihood (MV)	-17738.47	

Multicolinealidad

	Factor de inflación de la varianza (VIF)
--	---

**ESTIMACIÓN DE LA PROBABILIDAD DE POBREZA POR INGRESOS EN ECUADOR MEDIANTE UN MODELO LOGIT,
2022**

Jefe de Hogar 1.437

Zona Rural 1.008

Ingreso Per Cápita 1.016

Edad 1.456

Predicciones

Curva ROC 0.9964

Capacidad predictiva 0.9867

Muestra (personas) 286476

Probabilidad de ser pobre

Promedio 19.30%

Predicciones

Curva ROC

En este estudio se obtuvo una capacidad predictiva de 0.9867, el área bajo la curva de 0.9964 (ver Tabla 1). Además se expone la siguiente tabla de contingencia.

Tabla 2. Valores curva ROC.

		Valores actuales	
		0	1
Predicciones	0	229201	1810
	1	1998	53467

Gráficamente se tiene

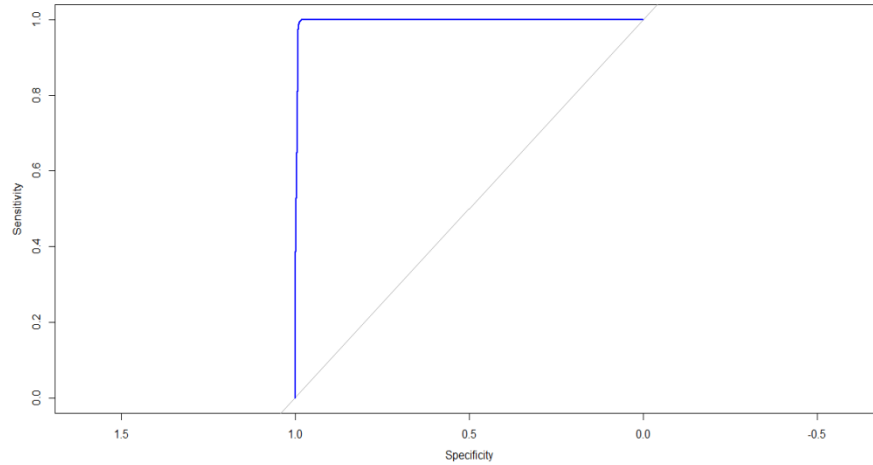


Figura 4. Curva ROC.

Predicciones del modelo logit muestral usando una muestra distinta

Para validar las predicciones del modelo logit muestral se usa la base de datos de 71620 observaciones por variable, correspondientes al 20% de la información obtenida de la base de datos del ENEMDU y el modelo de la pobreza en función las variables Jefe de Hogar, Zona Rural, Ingreso Per Cápita y Edad.

Se encontró una capacidad predictiva de 0.9868, el área bajo la curva de 0.9964 (ver Tabla 1). También, se evidencia la siguiente tabla de contingencia.

Tabla 3. Predicciones del modelo Logit.

Valores actuales

		0	1
Predicciones	0	57369	431
	1	512	13308

Se evidencia resultados prácticamente iguales a las predicciones de la primera base de datos cuando se usó el 80% de los mismos.

Modelo Logit poblacional

Parámetros

Como se observa en la tabla 5, en la estimación nacional (14427029 de personas), dos de las cuatro variables del modelo logit muestral son estadísticamente significativas, Ingreso Per Cápita y Edad.

Efectos Marginales

Ingreso Per cápita: Dado un incremento de un dólar en la renta de los individuos del Ecuador, la posibilidad de ser pobre se reduce en un 0.34%, manteniendo constante las demás variables.

Edad: Dado un aumento de un año en la edad de las personas que viven en el país, la probabilidad de ser pobre por ingresos baja en un 0.02%, con las demás variables constantes.

OddsRatios

Ingreso per cápita: Dado un incremento de un dólar en los ingresos de las personas ecuatorianas, la probabilidad de ser pobre se multiplica 0.798. Es decir, los ingresos per cápita están relacionados con una disminución de la posibilidad de ocurrencia del evento.

Edad: Por un aumento de un año en la edad de los individuos que viven en el país, la posibilidad de ser pobre se multiplica por 0.989. En este sentido, la Edad se relaciona con una disminución en la probabilidad de ocurrencia del suceso.

Tabla 4. Resultados del modelo logit poblacional.

Estimación				
$\widehat{Pobreza}_i = 19.800 - 0.225(\widehat{Ingreso\ per\ cápita}_i) - 0.011(\widehat{Edad}_i)$				
	Estimate	Std. Error	t value	Pr(> t)
Intercepto	19.800	0.74337	26.635	0.000
Jefe de Hogar	-0.067	0.048192	-1.395	0.163
Zona Rural	0.170	0.09347	1.817	0.069
Ingreso Per cápita	-0.225	0.008348	-26.969	0.000
Edad	-0.011	0.001594	-6.782	0.000
Efectos Marginales				

- El parámetro de la incógnita Zona Rural es positivo, indicando que las personas que viven en la ruralidad son más propensas a ser pobres, ya que no cuentan con fuentes de empleo que les proporcionen una remuneración adecuada, su educación es ineficiente, y los servicios de salud inadecuados (20).

Las personas que viven en el campo en el Ecuador, se sostienen en mayor medida de la agricultura no tecnificada y ganadería.

- El indicador de la variable Ingreso per Cápita es negativo en relación con la pobreza por ingresos; algo lógico ya que, a mayores ingresos o renta, menor probabilidad de ser pobre por ingresos en el Ecuador.
- El signo del estadístico de la variable Edad es negativo. Si una persona tiene mayor edad está en mejores capacidades de desempeñar un trabajo que un individuo de corta edad (9).

Efectos Marginales

Para encontrar los efectos marginales en un modelo logit se aplica la misma metodología que en una regresión lineal, es decir, derivar parcialmente cada variable e interpretar su resultado en forma porcentual. En este sentido, la interpretación de mencionadas derivadas nos muestra como un cambio en una variable predictora provoca un cambio en la probabilidad de ocurrencia del evento manteniendo las demás constantes (13).

Se realiza un análisis comparativo de los resultados de esta investigación y el estudio realizado en Montería-Colombia, donde se evidencia que los efectos marginales mantienen los mismos signos que los parámetros estimados, algo suficiente y necesario para este indicador. Si bien es ciertos en los dos hallazgos los resultados porcentuales son mínimos, sin embargo, son estadísticamente significativos para decir que son diferentes de cero (9).

OddsRatios

Los *odds* son la razón de probabilidad de que un evento ocurra, dividido para la probabilidad de que el suceso no ocurra (21).

$$odds_x = \frac{p_x}{1 - p_x} \quad (16)$$

En este sentido los *oddsRatios* (OR) es el cociente entre *odds*, dados dos valores diferentes de la variable independiente (x_1, x_2), muestra como la variación en los *odds* del suceso de interés por unidad de cambio en la variable predictora.

$$oddsRatios = \frac{\frac{p_{x_1}}{1 - p_{x_1}}}{\frac{p_{x_2}}{1 - p_{x_2}}} \quad (17)$$

Si el valor OR = 1, la variable explicativa no genera efecto en la probabilidad del suceso.

Si el valor $OR > 1$, la variable de entrada está ligada con un incremento en la probabilidad del evento.

Si el valor $OR < 1$, la variable predictora está relacionada con una disminución en la probabilidad del evento.

En consecuencia, los *oddRatios* de las variables independientes del modelo logit exponen que todas generan cambios en la probabilidad de que las personas sean pobres por ingresos en el Ecuador. Dichos cambios son correctos ya que en las variables Jefe de Hogar, Ingreso per Cápita, y Edad, las variaciones son negativas y en la incógnita Zona Rural los cambios son positivos.

Evaluación del Modelo

Significancia Individual: Para el nivel de significancia individual se utiliza la prueba z , donde todos los parámetros del modelo son estadísticamente significativos, es decir son diferentes de cero, por tal motivo todos los signos de las variables son representativos.

Significancia Global: La medida de significancia global de un modelo logit es el estadístico de la razón de verosimilitud (RV) que sigue una distribución χ^2 con grados de libertad igual al número de variables independientes. En consecuencia, podemos decir con certeza que el modelo en conjunto es estadísticamente significativo, proporcionándonos seguridad en la interpretación de los signos (12).

Bondad de Ajuste

Guevara y Uribe (13), consideran que en los modelos de variable dependiente binaria cualitativa, el valor de bondad de ajuste R^2 no es útil, en consecuencia recomiendan usar el *seudo* R^2 .

En la investigación se usa el $R^2 McFadden$ como medida de bondad de ajuste. Al igual que el R^2 varía entre 0 y 1 (12), su valor en el estudio es de 0.87. Es decir que el modelo explica el 87% de la variabilidad de las observaciones comparado con un modelo nulo.

Otra medida de bondad de ajuste que se usa es el Log Likelihood (MV) o logaritmo de la función de verosimilitud. Este estadístico muestra cuanta información no se explicó en la variable regresada luego del ajuste (21). En este sentido el modelo tiene una bondad de ajuste deseada.

Multilinealidad

Un paso previo a la presentación de un modelo logit es considerar si el mismo tiene problemas de multilinealidad (22).

Gujarati y Porter (12), dicen que la multilinealidad es un problema dentro de la regresión lineal donde las variables independientes están correlacionadas entre sí, generando grandes varianzas y covarianzas que dificultan la estimación de los parámetros, intervalos de confianza amplios influyendo que se acepte la hipótesis nula donde los parámetros son iguales a cero estadísticamente, entre otros inconvenientes.

Una medida que se utiliza para observar la velocidad de aumento de las varianzas y covarianzas es el factor inflacionario de la varianza (FIV), mostrando presencia de multicolinealidad.

$$FIV = \frac{1}{(1 - r^2)} \quad (18)$$

Estos autores consideran que un modelo econométrico tiene una alta multicolinealidad cuando su FIV es mayor a 10.

En el modelo muestral, todas las variables muestran un FIV menor de 1.5, exponiendo que no tiene inconvenientes de multicolinealidad, es decir que las variables predictoras no están correlacionadas.

Además, el logit muestral indica que la probabilidad promedio de ser pobre en el Ecuador es de 19.30%. Datos reales del INEC muestran que la pobreza por ingresos en el Ecuador en el año 2022 fue del 25,20 % aproximado. Esta diferencia podría explicarse porque se están omitiendo algunas variables importantes. Sin embargo, las probabilidades calculadas están cercanas al valor real.

Predicciones

Curva ROC: La curva ROC (Receiver Operating Characteristic) indica de mejor manera la capacidad predictiva de un modelo, brindando la potencia predictiva para el conjunto de valores referentes. Basada en la sensibilidad (posibilidad de predecir correctamente el éxito de la variable dependiente) y en la especificidad (probabilidad de predecir correctamente el fracaso de la variable respuesta), la curva ROC es un gráfico en el que se encuentra la sensibilidad en función de 1 menos la especificidad. Consecuentemente, si mayor es el área bajo la curva, mejores son las predicciones (21).

Matemáticamente tenemos:

$$\text{sensibilidad} = \frac{VP}{VP + FN}, \quad \text{especificidad} = \frac{VN}{VN + FP} \quad (19)$$

Donde, dichos valores son obtenidos de la siguiente tabla de contingencia

Tabla 5. Contingencia.

		Valores actuales	
Predicciones		Verdaderos Positivos (VP)	Falsos Positivos (FP)
		Falsos Negativos (FN)	Verdaderos Negativos (VN)

Fuente: (21).

En esta investigación se dividió la muestra en dos partes para demostrar su capacidad predictiva. Un 80% de los datos se destinaron para predecir el modelo, y el 20% de las observaciones restantes se usaron para comprobar que las predicciones sean semejantes o las mismas. Los resultados mostraron cifras muy semejantes en las dos muestras, una capacidad predictiva de 0.9867, el área bajo la curva de 0.9964. Consecuentemente se puede decir que el modelo tiene una gran capacidad predictiva.

Modelo Logit poblacional

Parámetros

El estudio se proyecta a nivel nacional utilizando el factor de expansión proporcionado por el INEC, interpretando al factor de expansión como el inverso de la probabilidad de elección de las unidades muestrales. Es así, que una unidad de la muestra representa a varias unidades poblacionales (8).

El logit poblacional indica que las variables Ingreso per cápita y Edad son estadísticamente significativas, es decir diferentes de cero, además los signos de las mismas son los esperados. Sin embargo, las variables Jefe de Hogar y Zona Rural quedan fuera del nuevo modelo poblacional. En este sentido, se demuestra que la renta de las personas y su edad son factores determinantes para explicar la pobreza en el país.

Efectos Marginales

Al igual que en el logit muestral, el modelo poblacional expone valores menores porcentuales de las variables Ingreso Percapita y Edad, pero estadísticamente significativos para decir para decir que influyen en las mismas.

OddsRatios

Los oddRatios de las incógnitas Ingreso per cápita y Edad del modelo logit poblacional, exhiben cambios en la probabilidad de que las personas sean pobres por ingresos en el Ecuador, dando variaciones negativas en las dos variables.

Finalmente, no es necesario hacer la evaluación del modelo, tampoco realizar la prueba de multicolinealidad y las predicciones, porque se está utilizando la totalidad de observaciones de la población.

CONCLUSIONES

Usando un modelo logit muestral y un modelo logit poblacional, el estudio calcula la probabilidad que tienen los individuos de llegar a la pobreza por ingresos, también muestra que factores conducen a las personas a este fenómeno social y estima predicciones de la misma, en el Ecuador en el año 2022.

En este sentido, el logit muestral determina la probabilidad de Pobreza por Ingresos de los individuos en el país, en función de cuatro variables explicativas: Jefe de Hogar, Zona Rural, Ingreso per Cápita y Edad. Además, mencionadas variables son los factores determinantes de la pobreza en nuestra sociedad, y finalmente dicho modelo presenta una gran capacidad predictiva.

Seguidamente, el logit poblacional también calcula la posibilidad de Pobreza por Ingresos de las personas en el Ecuador, pero con dos variables exógenas: Ingreso Per cápita y Edad. Por consiguiente, estas dos incógnitas son las determinantes de la pobreza. Finalmente en este modelo no es necesario hacer predicciones por se cuenta con el universo poblacional.

Es así, que se puede concluir diciendo que se cubrieron los objetivos planteados en la investigación y que la misma puede servir como base para nuevos estudios.

REFERENCIAS BIBLIOGRÁFICAS

1. Navarro LM, Ramos MAS. *Responsabilidad social y sostenibilidad: disrupción e innovación ante el cambio de época*. 2021.
2. Organización de las Naciones Unidas. *Objetivos de Desarrollo Sostenible: 17 objetivos para transformar el mundo* [Internet]. [citado s. f.]. Disponible en: <https://www.un.org/es/exhibits/17-objetivos-para-transformar-el-mundo>
3. Food and Agriculture Organization. *Objetivos de desarrollo sostenible*. Rome: FAO; 1986.
4. Instituto Nacional de Estadística y Censos (INEC). *Boletín de pobreza* [Internet]. Quito: INEC; 2022 jun [citado 2026 Mar 30]. Disponible en: https://www.ecuadorencifras.gob.ec/documentos/web-inec/POBREZA/2022/Junio_2022/202206_Boletin_pobreza.pdf
5. Instituto Nacional de Estadística y Censos (INEC). *Reporte de pobreza y desigualdad* [Internet]. Quito: INEC; 2022 dic [citado 2026 Mar 30]. Disponible en: https://www.ecuadorencifras.gob.ec/documentos/web-inec/POBREZA/2022/Diciembre_2022/202212_PobrezayDesigualdad.pdf
6. Instituto Nacional de Estadística y Censos (INEC). *Pobreza por ingresos: resultados históricos 2022* [Internet]. Quito: INEC; 2022 [citado 2026 Mar 30]. Disponible en: <https://www.ecuadorencifras.gob.ec/pobreza-ingresos-resultados-historico-2022/>
7. Instituto Nacional de Estadística y Censos (INEC). *Boletín técnico anual: Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU), pobreza y desigualdad (enero–diciembre 2022)* [Internet]. Quito: INEC; 2023 ene [citado 2026 Mar 30]. Disponible en: https://www.ecuadorencifras.gob.ec/documentos/web-inec/POBREZA/2022/Diciembre_2022/202212_PobrezayDesigualdad.pdf
8. Instituto Nacional de Estadística y Censos (INEC). *Diseño muestral: Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) anual 2022*. Quito: INEC; 2022.
9. Narváez ARA, Parra JB, Burgos LCV. *Modelo Probit para la medición de la pobreza en Montería, Colombia*. *Opción*. 2015;31(78):42–64.
10. Pucutay F. *Los modelos logit y probit en la investigación social*. 2001.
11. Marrugo-Arnedo CA. *Determinantes de la pobreza en la región Caribe colombiana*. *Rev Econ Caribe*. 2015;(15):47–69.

12. Gujarati DN, Porter DC. *Econometría*. 5a ed. México DF: McGraw-Hill; 2010.
13. Rodríguez Guevara DE, González Uribe GJ. *Principios de econometría*. Medellín: Instituto Tecnológico Metropolitano; 2017.
14. Segura EO, Torres V. Tratamiento de valores perdidos y atípicos en la aplicación del modelo estadístico de medición de impacto en un estudio de 90 fincas lecheras en la provincia de Pastaza, Ecuador. *Rev Cubana Cienc Agric*. 2014;48(4):333–6.
15. Zapata-Tapasco A, Pérez-Londoño S, Mora-Flórez J. Método basado en clasificadores k-NN parametrizados con algoritmos genéticos y la estimación de la reactancia para localización de fallas en sistemas de distribución. *Rev Fac Ing Univ Antioquia*. 2014;(70):220–32.
16. Batista G. *A study of k-nearest neighbour as an imputation method*. 2002.
17. Cubas Rovira GM. *Análisis del algoritmo MINI para imputación de valores perdidos en conjuntos de datos pequeños y con variables continuas y categóricas*. 2017.
18. Anderson DR, Sweeney DJ, Williams TA. *Estadística para administración y economía*. México: Cengage Learning; 2008.
19. Aranda VT. Generación de números aleatorios. *Suma*. 1995;(20):91–8.
20. Ray D. *Economía del desarrollo*. Barcelona: Antoni Bosch; 2022.
21. Ortuño M. Un modelo de regresión logística para el análisis de los aspectos que influyen en la anulación de pólizas de seguros de automóviles [Internet]. 2022 [citado 2026 Mar 30].
22. Gatica Neira F. Revisión empírica al modelo de desarrollo endógeno a partir de la innovación empresarial en Chile. *Innovar*. 2022;30(77):137–52.