



Doi: <https://doi.org/10.70577/asce.v5i3.1065>

Recibido: 2026-07-12

Aceptado: 2026-07-26

Publicado: 2026-08-11

**Minería de datos y econometría con modelos binarios y censurados:
determinantes de la exclusión del seguro médico en Ecuador**

**Data mining and econometrics with binary and censored models:
determinants of health insurance exclusion in Ecuador**

Autor

Luis Patricio Juna Pozo ¹

Facultad de Ciencias Económicas, Carrera de Ingeniería

lpjuna@uce.edu.ec

<https://orcid.org/0009-0005-9006-0502>

Universidad Central del Ecuador Estadística

Quito – Ecuador

Como Citar

Juna Pozo , L. P. (2026). Minería de datos y econometría con modelos binarios y censurados: determinantes de la exclusión del seguro médico en Ecuador. *ASCE MAGAZINE*, 5(3), 4042–4064.
<https://doi.org/10.70577/asce.v5i3.1065>



Resumen

En la presente investigación se analizó la exclusión del seguro médico en Ecuador como un problema de vulnerabilidad estructural asociado a la informalidad laboral. Aunque la tasa de empleo adecuado se situó en apenas un 34.7% al cierre de 2024, existió una brecha de conocimiento respecto a la cuantificación precisa de dicha desprotección mediante modelos para datos censurados y microdatos nacionales actualizados. Por ello, el objetivo consistió en identificar los determinantes socioeconómicos que explicaron la falta de afiliación al sistema de salud con base en la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) 2024. La metodología combinó técnicas de minería de datos y econometría avanzada bajo un enfoque cuantitativo explicativo. Se ejecutó un proceso de extracción, transformación y carga (ETL), seguido de una estimación comparativa entre el Modelo de Probabilidad Lineal, el modelo Probit y el modelo Tobit; con inclusión de factores de expansión muestral y pruebas de multicolinealidad. Los resultados, evaluados mediante efectos parciales promedio, evidenciaron que la ocupación informal constituyó el principal determinante de la exclusión, seguida por el área de residencia, el sexo y la condición de pobreza. Se concluyó que la convergencia entre minería de datos y modelos de respuesta limitada capturó con exactitud la intensidad de la vulnerabilidad, lo que aportó insumos técnicos para el diseño de políticas públicas orientadas a la formalización del empleo y a la cobertura efectiva de aseguramiento en el país.

Palabras clave: Exclusión del seguro médico, Protección social, Mercado laboral, Informalidad laboral, Modelos de respuesta binaria, Ecuador.



Abstract

In the present study, health insurance exclusion in Ecuador was analyzed as a structural vulnerability issue associated with labor informality. Although the adequate employment rate stood at barely 34.7% by the end of 2024, a knowledge gap persisted regarding the precise quantification of such social desprotection through models for censored data and updated national microdata. Therefore, the objective was to identify the socioeconomic determinants that explained the lack of affiliation to the healthcare system based on the 2024 National Employment, Unemployment, and Underemployment Survey (ENEMDU). The methodology combined data mining techniques and advanced econometrics under a quantitative explanatory approach. An extraction, transformation, and load (ETL) process was executed, followed by a comparative estimation among the Linear Probability Model, the Probit model, and the Tobit model; including sampling expansion factors and multicollinearity tests. The results, evaluated through average partial effects, evidenced that informal labor occupation constituted the main determinant of exclusion, followed by area of residence, sex, and poverty status. It was concluded that the convergence between data mining and limited dependent variable models captured the intensity of vulnerability with precision, which provided technical inputs for the design of public policies aimed at labor formalization and effective health insurance coverage in the country.

Keywords: Health insurance exclusion, social protection, Labor market, Informal employment, Binary response models, Ecuador.

Introducción

La seguridad en salud constituye un mecanismo institucional de protección frente a riesgos e incertidumbres que afectan la estabilidad económica de los hogares y la productividad nacional, en las economías en desarrollo, el acceso a sistemas de aseguramiento no solo garantiza el derecho a la salud, sino que funciona como un indicador determinante de la formalidad y la calidad del empleo (OIT, 2024; Cecchini et al., 2021). En el contexto ecuatoriano, a pesar de los esfuerzos normativos, las brechas de cobertura persisten de manera estructural, según el Instituto Nacional de Estadística y Censos (INEC 2025), la tasa de empleo adecuado se ubicó en apenas un 34.7% al cierre de 2024, lo que evidencia que la mayoría de la población activa permanece en condiciones de vulnerabilidad y exclusión del sistema de un seguro de salud de calidad (INEC, 2025; OIT, 2024).

Desde la perspectiva de la economía laboral, la probabilidad de que un individuo carezca de seguro médico está estrechamente vinculada a la teoría del capital humano y la segmentación del mercado, la literatura contemporánea sugiere que variables como el nivel de instrucción y la experiencia laboral no solo incrementan el ingreso, sino que actúan como barreras de entrada a sectores con mayor protección social (Scott y Freese, 2023; Silva et al., 2023). Por otra parte, factores demográficos y socioeconómicos exacerban estas disparidades; donde la ubicación geográfica y la condición de pobreza determinan una mayor propensión a la informalidad, esto a limitado el acceso a los beneficios del aseguramiento (Vélez, 2025; Banco Mundial, 2022).

El análisis de estos fenómenos requiere herramientas estadísticas que permitan modelar comportamientos dicotómicos de manera precisa, mientras que el Modelo de Probabilidad Lineal (MPL) ofrece una interpretación intuitiva, presenta limitaciones teóricas críticas como la heterocedasticidad y la posibilidad de generar predicciones fuera del intervalo lógico de cero a uno (Triola, 2022). Por ello, el uso de modelos de respuesta binaria, específicamente Probit y Tobit, resulta superior al capturar relaciones no lineales y asegurar que las probabilidades estimadas sean consistentes con la naturaleza discreta de los microdatos (Scott y Freese, 2023). Estos modelos permiten aislar con rigor el efecto de variables como la ocupación o el ingreso sobre la probabilidad de desprotección y no tener el rigor de validar problemas homocedásticos o de normalidad.

La aplicación de modelos de respuesta binaria para el estudio de microdatos provenientes de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU 2024) ha demostrado una alta eficacia en la identificación de patrones críticos de la realidad social ecuatoriana, al respecto, investigaciones previas sobre el bienestar poblacional han utilizado con éxito la regresión logística binaria o probit para modelar factores de riesgo como el sedentarismo, se pudo evidenciar que las variables sociodemográficas y laborales no solo condicionan la salud física, sino que estructuran las barreras de acceso a servicios fundamentales (Juna, Sedentarismo en Jóvenes y Adultos en Ecuador: Análisis con Ciencia de Datos y Regresión Logística Binaria (ENEMDU 2023) 2026). Este antecedente metodológico, sumado a la actualización de las cifras oficiales del Instituto Nacional de Estadística y Censos (INEC 2025), justifica la pertinencia de expandir el análisis hacia modelos Probit y Tobit, el uso de estos modelos econométricos permite una comprensión más profunda de la falta de adquisición de un seguro médico como un fenómeno de exclusión sistemática en el mercado laboral contemporáneo e incluso un problema de afiliación social obligatoria.

El marco del presente artículo tiene como objetivo analizar los determinantes socioeconómicos de la exclusión de seguro médico en Ecuador mediante el uso de la ENEMDU anual 2024, la investigación se justifica en la medida que el rol del investigador es contribuir a la solución de las complicaciones sociales, en este contexto, se pueden realizar posteriormente acciones administrativas denominada la toma de decisiones (Juna 2025); bajo esta premisa, surge la necesidad de actualizar la evidencia empírica en un entorno postpandemia donde las dinámicas laborales han mutado hacia una mayor precarización (OIT, 2024; CEPAL, 2023), a través de una comparación rigurosa entre modelos MPL, Probit y Tobit, se pretende identificar los factores con mayor peso estadístico en la exclusión del seguro médico, esto proporciona insumos técnicos para el diseño de políticas públicas orientadas a reducir la brecha de aseguramiento en los sectores más críticos de la economía nacional.

Material y métodos

La presente investigación se desarrolló bajo un enfoque cuantitativo, debido a que empleó técnicas estadísticas y econométricas para analizar la relación entre variables socioeconómicas y la exclusión del seguro médico en Ecuador, corresponde a una investigación de tipo aplicada, ya que

buscó generar evidencia empírica que contribuya a la formulación de políticas públicas orientadas a mejorar la cobertura del aseguramiento en salud. Su alcance fue explicativo, al identificar la asociación entre un conjunto de variables independientes y la probabilidad de no contar con seguro médico, el estudio adoptó un diseño no experimental, transversal y observacional, puesto que analizó información secundaria correspondiente a un único periodo de tiempo, sin manipular las variables objeto de estudio.

Como fuente de información se utilizaron los microdatos de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU), en su edición anual acumulada de 2024, publicada por el Instituto Nacional de Estadística y Censos (INEC, 2025). De acuerdo con el diseño muestral oficial, esta encuesta emplea un muestreo probabilístico, bietápico y por conglomerados, lo que garantiza representatividad nacional para las áreas urbana y rural, así como para los principales dominios provinciales, para el análisis se seleccionó la población económicamente activa de 15 años o más y se incorporaron los factores de expansión poblacional (fexp), con el propósito de obtener estimaciones representativas de la población ecuatoriana. Al tratarse de una base de datos oficial de acceso público, no se diseñó un instrumento propio de recolección de información ni fue necesaria su validación mediante juicio de expertos, el uso de los microdatos de la ENEMDU permitió analizar la exclusión del seguro médico a partir de información representativa del mercado laboral ecuatoriano (Florez y Aldana 2025).

Proceso ETL

Previo al análisis estadístico, se implementó un proceso de extracción, transformación y carga (ETL) sobre los archivos originales en el formato de las bases de datos que proporciona el INEC; en la fase de extracción, se integraron los microdatos provenientes de la encuesta para su procesamiento inicial, durante la fase de transformación, se realizó una depuración exhaustiva en el software IBM SPSS Statistics en su versión 26 (IBM Corp. 2019), lo que permitió la normalización de las variables, el tratamiento de datos atípicos y la validación de consistencia lógica, este proceso de limpieza resultó fundamental para garantizar la integridad de la información, especialmente en variables sensibles como los ingresos laborales y si tenían los participantes en la encuesta seguro médico, esta es la variable dependiente para los análisis econométricos y posteriormente cargar los datos en el entorno de desarrollo integrado RStudio,

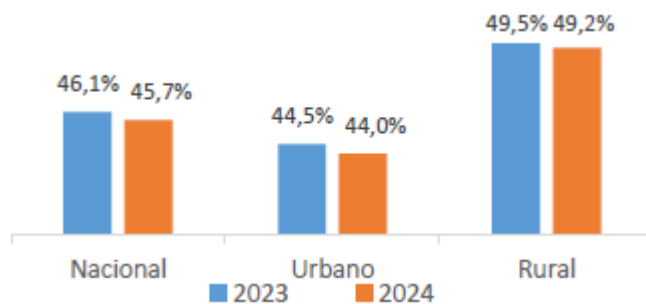
donde se estructuraron en objetos de datos optimizados para la ejecución de los modelos econométricos, contribuye para asegurar la consistencia interna de la muestra para las estimaciones finales.

Extracción

En la fase de extracción, se conservó la estructura original y garantizó la integridad de los datos, la precisión en el procesamiento de fuentes homogéneas resultó fundamental para evitar sesgos, errores de codificación y pérdidas de información en encuestas sociales (Alleva et al., 2021). Se seleccionó el conjunto de microdatos de la ENEMDU anual 2024 debido a sus ajustes metodológicos de comparabilidad en indicadores laborales, en la figura 1, la tasa de participación global alcanzó el 63,4% a nivel nacional, con un 61,5% en el área urbana y un 67,6% en el área rural, lo que aseguró la representatividad de la muestra frente a la dinámica económica vigente (INEC, 2025). Este proceso constituyó el primer filtro de calidad dentro del flujo de trabajo, el cual inició con la conexión a la fuente oficial y continuó con la validación de metadatos y etiquetas originales. Dicha revisión técnica confirmó que las variables primarias, tales como la ausencia de seguro médico, la condición de aseguramiento y el nivel de ingresos, conservaron la codificación requerida para su posterior tratamiento, la estabilidad en esta etapa garantizó que las variaciones en las estimaciones finales respondieran a la estructura del mercado laboral y no a distorsiones generadas durante la captura de la información.

Figura 1

Tasa de participación bruta a nivel nacional, urbano y rural, años 2023 y 2024



Elaborado por: Boletín Técnico N° 05-2025-ENEMDU.

Transformación

Durante la fase de transformación, se realizó una limpieza de los microdatos, con la finalidad de no tener datos perdidos o alguna inconsistencia a través del uso del software IBM SPSS Statistics en su versión 26 (IBM Corp, 2019), este proceso inició con la armonización de las variables originales de la ENEMDU 2024, esto integra la coherencia entre las etiquetas de los datos y los objetivos de la investigación, se aplicaron filtros de consistencia lógica para el tratamiento de valores perdidos y la identificación de valores atípicos en las variables de ingreso y horas trabajadas, lo que permitió mitigar sesgos en las estimaciones posteriores, la precisión en esta etapa de limpieza es fundamental para garantizar que el análisis de la precariedad no se vea alterado por errores de registro en la fuente primaria (Alleva et al., 2021).

Una parte crítica de la transformación fue la construcción de las variables operativas del modelo, la variable dependiente (VD) se operacionalizó bajo el concepto de no tener seguro médico esto por ser la variable censurada, es decir, que en su mayoría de participantes no poseen este beneficio social, donde el valor uno (1) representa la carencia de cobertura, por otro lado, se codificó la variable de ingresos económicos como referencia el Salario Básico Unificado (SBU) vigente en Ecuador para el año 2024, el cual se fijó en 460 dólares, esta variable fue dicotomizada con la asignación de uno (1) aquellos individuos cuyos ingresos laborales son mayores o iguales a 460 dólares, y el valor cero (0) a quienes perciben una remuneración inferior, esta segmentación permite identificar la relación entre la capacidad económica y el acceso al sistema de aseguramiento, basado en la toma de decisiones administrativas sobre el contexto laboral actual (Juna 2025).

El resto de variables categóricas se las recodificó con “1” para el nivel de factor que se desea analizar y “0” con el factor o el resto de factores de comparación, en este caso en las variables dicotómicas no requirió de mayores cambios.

Carga

En la fase de carga, los datos previamente depurados fueron incorporados al entorno analítico de RStudio, garantiza esto la persistencia estructurada de la información y su trazabilidad, un aspecto crítico en esta etapa fue el tratamiento de la variable dependiente relativa de ausencia de seguro médico, donde se aplicaron protocolos de validación para identificar registros incompletos, para

garantizar la potencia estadística del modelo, se procedió con la imputación de datos perdidos mediante métodos de consistencia lógica y asignación estocástica, con ello se evita sesgos de selección que pudieran subestimar la falta de cobertura en ciertos estratos, este proceso de gestión de la "no respuesta" es fundamental para mantener la representatividad de la muestra y asegurar que las inferencias sobre la precariedad laboral sean robustas y replicables (Little y Rubin 2022), finalmente, los datos fueron estructurados en objetos optimizados para la ejecución de los modelos MPL, Probit y Tobit.

Variables analizadas

La selección de las variables, se realizó mediante la relación o asociación de éstas por la correlación de Pearson y el estado del arte a lo que se refiere con los determinantes salariales (Tabla 1.).

Tabla 1
Variables usadas en los modelos econométricos

Variable	Tipo	Descripción	Codificación (ejemplo)
VD	Dicotómica	No recibe seguro médico	1 = No afiliado; 0 = Afiliado
area	Dicotómica	Área de residencia	1 = Urbana; 0 = Rural
Sexo	Dicotómica	Sexo del encuestado	1 = Mujer; 0 = Hombre
D_Edad	Dicotómica	Edad mayores de 25 años	1 = ≥ 25 años; 0 = < 25 años
D_Asal	Dicotómica	Asalariados con sueldo básico (460 USD)	1 = Sí; 0 = No
D_Ingr	Dicotómica	Ingresos ≥ 460 USD	1 = Sí; 0 = No
D_Horas	Dicotómica	Trabajo ≥ 40 horas semanales	1 = Sí; 0 = No
D_Civil	Dicotómica	Estado civil	1 = Casado; 0 = Otros
D_Instrucción	Dicotómica	Educación media	1 = Secundaria; 0 = Otros niveles
D_Etnia	Dicotómica	Identificación étnica (mestizos)	1 = Mestizo; 0 = Otros
D_Ocupación	Dicotómica	Tipo de ocupación (trabajo dependiente)	1 = Dependiente; 0 = Otros
D_Lugar	Dicotómica	Sitio donde trabaja o produce	1 = Formal; 0 = Informal
D_Exper	Dicotómica	Experiencia laboral ≥ 5 años	1 = Sí; 0 = No
D_Parentezco	Dicotómica	Posición en el hogar	1 = Jefe de hogar; 0 = Otros
D_NumTrab	Numérica	Número de trabajos	1 = uno; 0 = más de uno
D_SeguroAc	Dicotómica	Aporta algún seguro actual	1 = Sí; 0 = No
fexp	Escala	Factor de expansión (ponderador muestral)	Valor numérico continuo

El dataset ENEMDU 2024 fue extraído de la página oficial del Instituto Nacional de Estadística y Censos (INEC), evidencia que la mayoría de los participantes no recibía ningún seguro médico; esta presencia de datos perdidos y la naturaleza censurada de la variable dependiente justificó el uso del modelo Tobit, el cual resulta metodológicamente superior en comparación con el Modelo de Probabilidad Lineal y el modelo Probit al capturar la censura en las observaciones y mejorar la validez de las estimaciones (Rahmani y Benmoumen, 2025; Kaur, 2025). De forma análoga, el resto de las variables independientes con valores perdidos fue transformado en regresores dicotómicos o dummy para su estandarización y representatividad, lo que otorgó mayor consistencia técnica al procesamiento en el software RStudio y garantizó la robustez de los modelos econométricos estimados (Angrist y Pischke, 2024; Cameron y Trivedi, 2005).

Modelos econométricos

La arquitectura analítica de esta investigación se fundamenta en la econometría de respuesta cualitativa, esencial cuando la variable dependiente es de naturaleza binaria, en el estudio de las brechas de ingresos, este enfoque permite modelar la probabilidad de que un individuo se encuentre en situación de desprotección en salud en función de un conjunto de predictores socioeconómicos, la literatura académica reciente destaca que el uso de microdatos provenientes de encuestas de hogares requiere un contraste riguroso entre especificaciones lineales y funciones de enlace no lineales para garantizar la validez de las inferencias y la precisión de los efectos marginales (Wooldridge, 2023; Scott y Freese, 2023), bajo esta premisa, se procedió a la estimación comparativa de tres estructuras estocásticas: el Modelo de Probabilidad Lineal (MPL), el modelo Logit y el modelo Probit.

Modelo de Probabilidad Lineal (MPL)

El Modelo de Probabilidad Lineal constituye la aproximación inicial para el estudio de exclusión de seguro médico, caracterizándose por tratar la variable dependiente binaria mediante una regresión de mínimos cuadrados ordinarios (MCO), esta especificación permite que los parámetros estimados se interpreten directamente como el cambio en la probabilidad de percibir ingresos inferiores al umbral establecido ante variaciones unitarias en las variables explicativas, aunque este modelo presenta desafíos teóricos, como la heterocedasticidad inherente y la posibilidad de obtener valores fuera del rango probabilístico estándar, la literatura econométrica actual lo valida como un

estimador consistente y de alta utilidad para capturar tendencias promedio en grandes conjuntos de datos (Wooldridge, 2023; Angrist y Pischke, 2021).

$$P(Y_i = 1|X_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i \quad (1)$$

Donde:

- Y_i es la variable dependiente binaria (no tener seguro médico)
- β_0 es la constante o intercepto
- β_j son los coeficientes a estimar
- ε_i es el término de error.

La lectura de los coeficientes en un modelo de probabilidad lineal es sencilla: cada vez que una variable independiente X_j aumenta en una unidad, la probabilidad de que el evento $Y=1$ se modifique en β_j , siempre que las demás variables permanezcan constantes, este principio es conocido como *ceteris paribus*, permite aislar el efecto específico de cada factor dentro del modelo y facilita la comprensión de cómo influyen las distintas características en el resultado observado (Wooldridge 2023).

Modelo Probit

La implementación del modelo Probit es un paso analítico crucial, ya que comparte la misma fundamentación teórica que el modelo Tobit en lo que respecta a la estructura de variables latentes, mientras que el Probit se limita a modelar la probabilidad dicotómica de estar desprotegido, el modelo Tobit extiende este análisis al considerar la intensidad del fenómeno en datos censurados, según Wooldridge (2023), el Probit es el componente que determina el mecanismo de selección (la probabilidad de no tener seguro médico sea mayor a cero o supere un umbral), mientras que la estructura de distribución normal subyacente permite que el modelo Tobit estime correctamente los coeficientes sin los sesgos que presentaría una regresión lineal truncada.

De esta manera, el Probit no solo ofrece una estimación de probabilidad, sino que valida el supuesto de normalidad necesario para modelos de respuesta limitada más complejos (Gujarati y Porter 2022). La especificación formal del modelo Probit se define mediante la probabilidad de que la

variable dependiente tome el valor de uno ($Y=1$), condicionada a un conjunto de variables explicativas X , matemáticamente, se expresa como:

$$P(Y = 1|X) = \Phi(X^T \beta) \quad (2)$$

Donde:

- $P(Y=1|X)$: s la probabilidad de que el evento ocurra (persona no tenga seguro médico).
- Φ representa la función de distribución acumulada normal estándar.
- $X^T \beta$: es el producto escalar entre el vector de variables independientes X y el vector de coeficientes β .

Este modelo es útil cuando se asume una distribución normal de los errores (Gujarati y Porter 2022).

Modelo Tobit (Regresión Censurada)

El modelo Tobit surge como una extensión necesaria cuando la variable dependiente no es puramente dicotómica, sino continua pero limitada o "censurada" en un umbral específico (comúnmente en cero), a diferencia del Probit, que solo analiza la probabilidad de no estar cubierto por un seguro de salud, el Tobit permite cuantificar la magnitud de esa desprotección, basándose en la misma premisa de normalidad de los errores que el Probit, el modelo se define a través de una variable latente Y^* de la siguiente manera:

$$Y_i^* = X_i \beta + \varepsilon_i, \varepsilon_i \sim \varepsilon_i(0, \sigma^2) \quad (3)$$

Donde la variable observada Y_i (por ejemplo, el nivel de ingresos por debajo del salario básico) se expresa como:

$$Y_i = \begin{cases} Y_i^* & \text{si } Y_i^* > 0 \\ 0 & \text{si } Y_i^* \leq 0 \end{cases} \quad (4)$$

Para garantizar una comparación válida entre los modelos de respuesta binaria y el modelo Tobit, se aplicó un procedimiento de ajuste basado en la literatura de datos censurados, dado que los coeficientes brutos del modelo Tobit no son directamente comparables con los del Modelo de Probabilidad Lineal (MPL), se utilizó un factor de suavizamiento que pondera los estimadores por la probabilidad de que la variable dependiente sea mayor a cero, este ajuste, fundamentado en la distribución normal acumulada, permite transformar los parámetros del Tobit en efectos marginales

sobre la media observada, siendo amigable para una interpretación coherente con los resultados obtenidos en el MPL, Logit y Probit bajo el principio *ceteris paribus* (Wooldridge 2023; Gujarati y Porter 2022).

Validación y pruebas estadísticas

Correlación de las variables

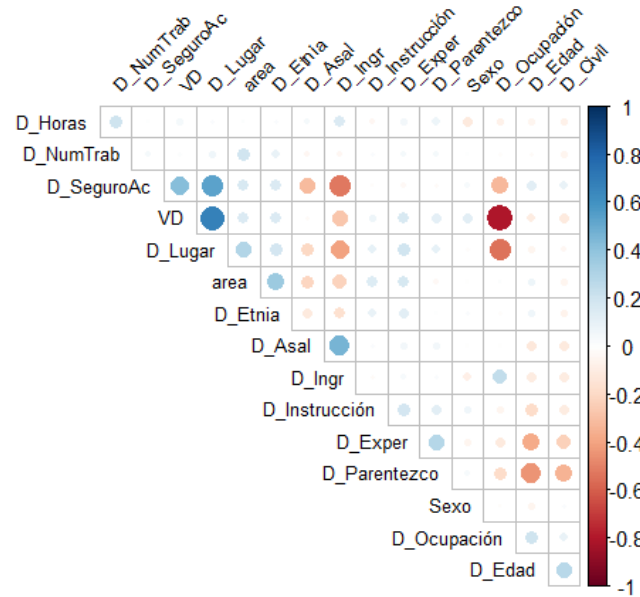
El análisis de correlación es una fase esencial en la econometría, pues identifica la dirección y fuerza de asociación entre variables previo a la estimación de modelos complejos (Triola 2022). en el estudio de la desprotección social, este diagnóstico permite validar la lógica de los determinantes seleccionados y detectar problemas de multicolinealidad, asegura que los modelos MPL, Probit y Tobit se sustenten en evidencia empírica coherente (Wooldridge 2023).

Las correlaciones que muestra la figura 2, la relación entre la variable dependiente (no tener seguro médico) y los determinantes socioeconómicos considerados en el modelo, el resultado más relevante corresponde a la ocupación ($D_{\text{Ocupación}} = -0.8243$) presenta la asociación negativa más fuerte, lo que sugiere que el empleo dependiente o formal reduce significativamente la probabilidad de exclusión del aseguramiento, también se observan relaciones negativas en los ingresos ($D_{\text{Ingr}} = -0.2748$), la edad mayor de 25 años ($D_{\text{Edad}} = -0.1131$) y el estado civil ($D_{\text{Civil}} = -0.1188$), lo que indica que mayores ingresos, mayor edad y estar casado tienden a disminuir la probabilidad de no contar con seguro médico. A diferencia de algunas variables presentan asociaciones positivas con la variable dependiente, se destaca que el lugar de trabajo ($D_{\text{Lugar}} = 0.7307$), muestra una relación fuerte y sugiere que determinadas condiciones laborales incrementan la vulnerabilidad frente a la desprotección social, mientras que aportar a algún seguro ($D_{\text{SeguroAc}} = 0.4308$) evidencia una asociación positiva moderada.

Otras variables como etnia, experiencia laboral, parentesco y número de trabajos presentan relaciones positivas de menor magnitud, lo que sugiere una influencia más limitada dentro de la estructura del modelo, evidencia que los factores vinculados al mercado laboral constituyen los determinantes más relevantes en la probabilidad de no tener seguro médico.

Figura 2

Matriz de correlación de los determinantes de la desprotección social en el Ecuador



Prueba de no multicolinealidad

La prueba de multicolinealidad es fundamental para garantizar la validez del artículo, ya que permite identificar si existe una relación lineal exacta o altamente aproximada entre las variables independientes que pueda sesgar los estimadores, la decisión sobre la validez de una variable se define mediante el Factor de Inflación de la Varianza (VIF); de acuerdo con Wooldridge (2023), valores de VIF superiores a 10 indican una colinealidad severa que requiere la eliminación o reestructuración de la variable para evitar varianzas infladas en los modelos MPL, Probit y Tobit.

En este estudio, los resultados del test VIF para el modelo muestran valores que oscilan entre 1.04 y 2.09, lo que confirma la ausencia de multicolinealidad y valida la inclusión de todos los regresores, en la figura 3, se confirma que los valores de VIF se mantienen significativamente por debajo del umbral crítico con un $VIF < 5$, se asegura que las intensidades observadas en la matriz de correlación no son producto de redundancias estadísticas, esto permite una interpretación robusta de los signos y magnitudes.

Figura 3

Prueba de multicolinealidad (VIF) para las variables del modelo econométrico

```

--- PRUEBA DE MULTICOLINEALIDAD (VIF) ---
> print(vif(mco_w))
      area      Sexo      D_Edad      D_Asal      D_Ingr
1.393528  1.042702  1.501391  1.354172  1.769741
  D_Horas  D_Civil D_Instrucción  D_Etnia  D_Ocupación
1.118180  1.208955  1.090400  1.191381  1.622796
  D_Lugar  D_Exper D_Parentezco D_NumTrab  D_SeguroAc
2.090057  1.328044  1.378231  1.090732  1.760575

```

Resultados

Estimación del Modelo de Probabilidad Lineal (MPL) y Validación de Consistencia.

Los resultados de la tabla 2, revelan un ajuste sólido con un coeficiente de determinación R^2 de 0.7781, sugiriendo que las variables seleccionadas capturan la mayor parte de la dinámica de la vulnerabilidad en el Ecuador, la significancia global del modelo ($p < 0.01$) y la magnitud del coeficiente de ocupación laboral (-0.668) confirman que la estabilidad en el empleo es el principal mecanismo de salida de la desprotección (Heckman 2001), el lugar de trabajo (0.294) y el sexo (0.134) presentan efectos positivos, señal que el sitio donde desarrolla su actividad laboral y ser mujer aumentan la probabilidad de no contar con seguro, en contraste, estar casado (-0.071) y tener más de un trabajo (-0.057) disminuyen la exclusión.

Tabla 1

Estimación del MPL para los determinantes de la desprotección social

Variable	Estimate	Std. Error	t value	Pr(> t)	Significancia
(Intercept)	0.412444	0.003791	108.789	< 2e-16	***
area	0.056757	0.001512	37.531	< 2e-16	***
Sexo	0.134122	0.001255	106.877	< 2e-16	***
D_Edad	0.078155	0.002037	38.359	< 2e-16	***
D_Asal	0.057624	0.001467	39.271	< 2e-16	***
D_Ingr	0.046584	0.001643	28.349	< 2e-16	***
D_Horas	0.005854	0.001508	3.881	0.000104	***
D_Civil	-0.071867	0.001424	-50.467	< 2e-16	***
D_Instrucción	-0.013364	0.001269	-10.532	< 2e-16	***
D_Etnia	0.073612	0.001674	43.961	< 2e-16	***
D_Ocupación	-0.668867	0.001585	-421.942	< 2e-16	***
D_Lugar	0.294817	0.001872	157.526	< 2e-16	***
D_Exper	0.043637	0.001421	30.714	< 2e-16	***
D_Parentezco	-0.061680	0.001677	-36.787	< 2e-16	***

D_NumTrab	-0.057154	0.002451	-23.320	< 2e-16	***
D_SeguroAc	0.095787	0.001729	55.394	< 2e-16	***

Los coeficientes estimados representan la asociación estadística entre cada variable independiente y la variable dependiente, manteniendo constantes las demás variables incluidas en el modelo, en este contexto, las variables área ($\beta = 0.0568$), sexo ($\beta = 0.1341$), edad ($\beta = 0.0782$), condición de asalariado ($\beta = 0.0576$), ingreso ($\beta = 0.0466$), horas trabajadas ($\beta = 0.0059$), etnia ($\beta = 0.0736$), lugar de residencia ($\beta = 0.2948$), experiencia laboral ($\beta = 0.0436$) y afiliación al seguro de salud ($\beta = 0.0958$) presentan coeficientes positivos y estadísticamente significativos ($p < 0.001$), estos resultados indican que, manteniendo constantes las demás variables del modelo, dichas características se asocian con un mayor valor esperado de la variable dependiente.

En contraste, las variables estado civil ($\beta = -0.0719$), nivel de instrucción ($\beta = -0.0134$), ocupación ($\beta = -0.6689$), parentesco ($\beta = -0.0617$) y número de trabajadores ($\beta = -0.0572$) presentan coeficientes negativos y estadísticamente significativos ($p < 0.001$), esto sugiere que, controlando el efecto de las demás covariables, estas variables se asocian con una reducción en el valor esperado de la variable respuesta.

Al comparar la magnitud de los coeficientes, la variable ocupación registra la asociación negativa más pronunciada ($\beta = -0.6689$), mientras que, el lugar de residencia presenta la asociación positiva de mayor magnitud ($\beta = 0.2948$); no obstante, la interpretación de estos coeficientes debe realizarse considerando la escala de medición y la codificación utilizada para cada variable, especialmente en el caso de variables dicotómicas o categóricas.

Evaluación Comparativa y Efectos Marginales: Probit y Tobit Ajustado

Los resultados de la tabla 2, muestran que las variables relacionadas con las condiciones laborales presentan las asociaciones estadísticas más relevantes con la exclusión del seguro médico, la ocupación registra la mayor magnitud de asociación en los modelos MPL y Tobit, seguida por el lugar de trabajo, el sexo, la edad, la condición de asalariado y el nivel de ingreso, cuyos coeficientes positivos indican una asociación con una mayor probabilidad de exclusión, en cambio el estado civil y el parentesco presentan coeficientes negativos, lo que se asocia con una menor probabilidad de exclusión respecto a las categorías de referencia.

Por otra parte, variables como instrucción, etnia, número de trabajos y seguro actual presentan diferencias en el signo de sus coeficientes entre los modelos, por lo que su interpretación debe realizarse con cautela, los resultados evidencian asociaciones estadísticamente significativas entre

los factores sociodemográficos, laborales y socioeconómicos y la exclusión del seguro médico, sin que ello implique relaciones de causalidad.

Tabla 2

Comparativa de efectos parciales en los modelos MPL, Probit y Tobit

Variable	MPL (B)	Probit AME	Tobit Adj	Análisis interpretativo
Intercepto	0.4124***	—	0.1093***	Probabilidad base de exclusión bajo condiciones de referencia.
area	0.0568***	0.0287***	0.0684***	Brechas territoriales: vivir en área urbana aumenta la exclusión.
Sexo	0.1341***	0.0463***	0.1445***	Las mujeres enfrentan mayor exclusión, refleja desigualdad estructural.
D_Edad	0.0782***	0.0324***	0.0863***	Grupos mayores de 25 años muestran mayor riesgo de exclusión.
D_Asal	0.0576***	0.0402***	0.0742***	Ser asalariado no garantiza cobertura; refleja precariedad laboral.
D_Ingr	0.0466***	0.0226***	0.0531***	Bajos ingresos refuerzan la exclusión del aseguramiento.
D_Horas	0.0059***	0.0274***	0.0031***	Jornadas laborales largas no aseguran afiliación; vinculadas al subempleo.
D_Civil	-0.0718**	-0.0151***	-0.0167***	El matrimonio reduce la exclusión, posiblemente por acceso vía cónyuge.
D_Instrucción	-0.0134***	0.0151***	0.0167***	La educación media protege frente a la exclusión.
D_Etnia	0.0736***	-0.0027**	0.0873***	Condiciones étnicas reflejan desigualdades estructurales.
D_Ocupación	-0.6688***	-0.2057***	-1.7485***	Determinante más fuerte: la informalidad laboral multiplica la exclusión.
D_Lugar	0.2948***	-0.029***	0.1816***	El sitio de trabajo informal incrementa notablemente la vulnerabilidad.
D_Exper	0.0043***	0.0043***	0.0093***	La experiencia laboral no asegura cobertura; persiste la exclusión.
D_Parentezco	-0.0617***	-0.0445***	-0.0587***	Ser jefe de hogar reduce la exclusión, por necesidad de protección familiar.
D_NumTrab	-0.0572***	0.063***	-0.0718***	Tener más de un empleo aumenta la probabilidad de contar con seguro en al menos uno.
D_SeguroAc	0.0958***	0.0895***	0.1006***	Tener algún seguro actual no garantiza calidad ni cobertura suficiente.

Los coeficientes estimados muestran una clara consistencia en la dirección de los efectos, lo que refuerza la robustez del análisis, los resultados evidencian que la exclusión del seguro médico no depende de un solo factor, sino de un entramado de condiciones laborales y sociodemográficas que configuran un patrón estructural de desprotección en el país.

Discusión

El objetivo de esta investigación consistió en identificar los determinantes socioeconómicos asociados con la exclusión del seguro médico en Ecuador mediante la aplicación de modelos econométricos para variables dependientes binarias sobre los microdatos de la ENEMDU 2024, los resultados evidenciaron que las características laborales, económicas y demográficas presentaron asociaciones estadísticamente significativas con la probabilidad de no disponer de cobertura de seguro médico, lo que confirmó que el acceso al aseguramiento respondió a un fenómeno multidimensional. Estos hallazgos coincidieron con la literatura económica que sostiene que las decisiones relacionadas con variables dicotómicas requieren modelos de respuesta limitada para estimar adecuadamente la probabilidad de ocurrencia del evento de interés, evitando las limitaciones inherentes a los modelos lineales convencionales (Greene, 2018; Wooldridge, 2019; Stock & Watson, 2020).

Las asociaciones observadas entre la condición laboral, la ocupación y la probabilidad de exclusión del seguro médico respaldaron la evidencia empírica que relacionó la inserción laboral con el acceso a mecanismos de protección social, de forma particular, la mayor vulnerabilidad de determinados grupos ocupacionales sugirió que la estabilidad del empleo continuó siendo un factor relevante para la afiliación a sistemas de aseguramiento. Esta evidencia resultó consistente con los planteamientos de Rahmani et al. (2023), quienes identificaron que la informalidad y las diferencias en las condiciones de empleo incrementaron las brechas de acceso a los servicios de salud y redujeron la probabilidad de contar con mecanismos formales de protección, los resultados del presente estudio reforzaron la necesidad de considerar el mercado laboral como uno de los principales determinantes del aseguramiento en salud.

Las variables sociodemográficas también presentaron asociaciones significativas con la exclusión del seguro médico, la evidencia mostró que factores como la edad, el sexo y el estado civil contribuyeron a explicar diferencias en la probabilidad de disponer de cobertura, resultado que coincidió con investigaciones recientes que identificaron desigualdades persistentes en el acceso a los sistemas de salud entre distintos grupos poblacionales, en este contexto, Kaur et al. (2022) señalaron que las condiciones socioeconómicas y demográficas modificaron la utilización de los servicios sanitarios y la incorporación a esquemas de aseguramiento, especialmente en países con elevados niveles de empleo informal como es el caso ecuatoriano, aunque las magnitudes de los efectos variaron entre estudios debido a diferencias institucionales y metodológicas, la dirección de las asociaciones mostró un comportamiento similar al observado en la presente investigación.

Desde el punto de vista metodológico, la consistencia en el signo y la significancia estadística de los coeficientes fortaleció la interpretación de las asociaciones identificadas entre las variables explicativas y la exclusión del seguro médico, este comportamiento coincidió con las recomendaciones formuladas por Maddala (1983), Greene (2018) y Wooldridge (2019), quienes señalaron que los modelos de respuesta limitada permiten obtener estimaciones consistentes cuando la variable dependiente adopta únicamente dos categorías mutuamente excluyentes. De igual manera, Stock y Watson (2020) indicaron que la interpretación de estos modelos debe centrarse en las asociaciones estadísticas y en las probabilidades estimadas, sin atribuir relaciones causales cuando la evidencia procede de estudios observacionales de corte transversal, en este sentido, los resultados obtenidos respaldaron el enfoque metodológico adoptado y proporcionaron evidencia robusta para el análisis de los factores asociados con la exclusión del seguro médico.

Los hallazgos permitieron cumplir el objetivo de la investigación al identificar factores socioeconómicos vinculados con la exclusión del seguro médico en Ecuador, la evidencia mostró que las diferencias en condiciones laborales, económicas y demográficas siguen condicionando el acceso al aseguramiento, lo que coincide con estudios que destacan la informalidad laboral como el principal obstáculo para la afiliación a la seguridad social (Sarmiento Moreno 2025). Las estrategias de cobertura deben incluir políticas que reduzcan la informalidad, fortalezcan el empleo formal y disminuyan las desigualdades socioeconómicas, en línea con investigaciones que señalan la necesidad de fortalecer la institucionalidad y la calidad del empleo para ampliar la protección social (Cupacán Lagos, 2026; Vargas Miño, 2026). El diseño transversal del estudio permitió asociaciones estadísticas, por lo que futuras investigaciones deberían incorporar datos

longitudinales o métodos cuasiexperimentales que permitan evaluar la evolución temporal de estos determinantes y robustecer la evidencia sobre el acceso al aseguramiento en salud (Mendoza-Falcones et al., 2026).

Conclusiones

Las conclusiones del estudio permiten afirmar que la exclusión en el acceso al seguro médico en Ecuador responde a determinantes estructurales del mercado laboral, el factor más influyente fue la ocupación: la informalidad laboral se consolida como el núcleo explicativo de la desprotección social, la falta de estabilidad contractual, la ausencia de afiliación obligatoria y la precarización del empleo limitan de manera significativa la cobertura en salud, de forma complementaria, el lugar de trabajo donde se desarrolla las actividades de los trabajadores ecuatorianos y las condiciones contractuales refuerzan este patrón, se mostró que la segmentación territorial y productiva inciden directamente en la probabilidad de no contar con protección social, de la misma manera, se evidencian desigualdades sociodemográficas persistentes: las mujeres presentan mayor probabilidad de exclusión, lo que refleja que las brechas de género en el mercado laboral se extienden también al acceso a derechos sociales básicos.

Desde el punto de vista metodológico, la consistencia en la dirección de los efectos entre los modelos MPL, Probit y Tobit fortalece la validez interna del análisis, el modelo Tobit, como se esperaba, aporta una lectura más robusta al capturar la censura inherente en la variable dependiente y reflejar la intensidad real de la exclusión, mientras que el Probit permite interpretar efectos marginales promedio con mayor precisión probabilística, esta convergencia confirma que la exclusión del seguro médico no es un fenómeno aislado, sino un proceso sistemático que se profundiza en contextos de informalidad y desigualdad estructural.

En consecuencia, los hallazgos sustentan la necesidad de políticas públicas orientadas a la formalización laboral, la reducción de brechas territoriales y la inclusión efectiva de grupos vulnerables en el sistema de protección social, adicionalmente, se abre una oportunidad para el sector privado de seguros médicos, que enfrenta el reto de diseñar productos más accesibles y confiables para los consumidores, esto contribuye a disminuir la exclusión y fortalecer la cobertura en salud en el país.



La evidencia obtenida también sugiere que la exclusión del aseguramiento médico debe analizarse en relación con la transformación digital y la innovación en los sistemas de protección social tanto en los sectores urbanos y rurales, la incorporación de tecnologías de información, registros electrónicos y plataformas de afiliación en línea puede facilitar el acceso, reducir costos administrativos para ampliar la cobertura con una mayor calidad de los servicios, especialmente en zonas rurales de difícil acceso y entre grupos tradicionalmente marginados, este enfoque complementa las políticas de formalización laboral y permite avanzar hacia un sistema más inclusivo, eficiente y sostenible.

Referencias Bibliográficas

- Alleva, Giorgio, Piero Falorsi, Paolo Righi, y Francesca Petrarca. «Measuring the Accuracy of Aggregates Computed from a Statistical Register.» *The Journal of Official Statistics (JOS)* 37, n° 2 (2021): 481-503.
- Angrist, Joshua D., y Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press, 2024.
- Banco Mundial. *The World Bank's social protection and labor strategy : resilience, equity, and opportunity - executive summary (Chino)*. Washington, DC: Banco Mundial, 2022.
- Cameron, A. Colin, y Pravin K. Trivedi. *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press, 2005.
- Cecchini, Simone, Raúl Holz, y Humbert Soto. *Sistemas de protección social en América Latina y el Caribe: una perspectiva comparada*. Santiago de Chile, 2021.
- CEPAL. «Panorama Social de América Latina y el Caribe 2022: la transformación de la educación como base para el desarrollo sostenible.» (CEPAL) 2023: 282.
- Cupacán Lagos, E. F. *Análisis de los factores socioeconómicos asociados a la informalidad laboral y afiliación al Instituto Ecuatoriano de Seguridad Social*. Ibarra: Universidad Técnica del Norte, 2026.



-
- Florez-Vaquiro, Nelson, y Lia Alejandra Hincapié-Aldana. «Precariedad laboral y juventudes: los mercados de trabajo en Argentina, Brasil, Colombia y México.» *Íconos - Revista de Ciencias Sociales*, n° 81 (2025): 13-32.
- Greene, William H. *Econometric Analysis*. Pearson, 2018.
- Gujarati, Damodar N., y Dawn C. Porter. *Essentials of Econometrics*. 6ta. Editado por SAGE Publications. New York / México D.F.: McGraw-Hill, 2022.
- Heckman, James J. «Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture.» *Journal of Political Economy* (The University of Chicago Press) 109, n° 4 (2001): 673-748.
- IBM Corp. «IBM SPSS Statistics para Windows.» 2019.
- INEC. «Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU), Anual 2024. Boletín Técnico N° 05-2025-ENEMDU.» (INEC) 2025.
- Juna, Luis. «Importancia de la investigación científica en los profesionales quiteños y el gusto en su profesión analizado por la prueba de concordancia Kappa de Cohen.» *Ciencia Latina Revista Científica Multidisciplinar* 9, n° 2 (2025).
- Juna, Luis. «Sedentarismo en Jóvenes y Adultos en Ecuador: Análisis con Ciencia de Datos y Regresión Logística Binaria (ENEMDU 2023).» *Technology Rain Journal* 5, n° 1 (2026).
- Kaur, Gursharan. «Using the Tobit model in social science research: Applications and insights.» *International Journal of Multidisciplinary Research Review* 11, n° 6 (2025).
- Little, Roderick J. A., y Donald B. Rubin. *Statistical Analysis with Missing Data 3ra edición*. Hoboken, NJ: John Wiley & Sons, 2022.
- Maddala, G. S., y Kajal Lahiri. *Introduction to Econometrics*. Hoboken, NJ: John Wiley & Sons, 2021.
- Mendoza-Falcones, L. M., G. N. Guida-Acevedo, C. A. Herrera-Esquivel, y P. J. Herrera-Sánchez. «Bioética y equidad en el acceso a los servicios de salud en Ecuador y América Latina.» *Journal of Economic and Social Science Research* 6, n° 1 (2026).
- OIT. «Perspectivas Sociales y del Empleo en el Mundo: Tendencias 2024.» Ginebra, 2024, 111.
- Rahmani, El Ouaid, y Mohamed Benmoumen. «The detection method of the Tobit model in a dataset.» *Stats* 8, n° 3 (2025): 59.
- Sarmiento Moreno, M. B. *Brechas de salud y de condiciones laborales en trabajo informal en la población de Ecuador*. Barcelona, España: Universitat Politècnica de Catalunya, 2025.



- Silva Filho, Telmo de Menezes, Hao Song, y Meelis Kull. «Classifier Calibration: A Survey on How to Assess and Improve Predicted Class Probabilities.» *Machine Learning* 112 (2023): 3211-3260.
- Stock, James H., y Mark W. Watson. *Introduction to Econometrics*. Pearson, 2020.
- Triola, Mario F. *Elementary Statistics*. Boston: Pearson Education, 2022.
- Vargas Miño, A. R. *Efecto del crecimiento económico, la complejidad económica, la informalidad y la calidad institucional sobre la afiliación a la seguridad social en Ecuador*. Loja: Universidad Nacional de Loja, 2026.
- Vélez, Viviana, Jessica Quispillo, y Ercilia Franco. «Informalidad laboral en la población económicamente activa de la ciudad de Guayaquil: Causas y Consecuencias.» *Revista La Universidad Segunda Época* 6 (2025).
- Wooldridge, Jeffrey M. *Introductory Econometrics: A Modern Approach*. 7ma. Boston: Cengage Learning, 2023.

Conflicto de intereses:

Los autores declaran que no existe conflicto de interés posible.

Financiamiento:

No existió asistencia financiera de partes externas al presente artículo.

Agradecimiento:**Nota:**

El artículo no es producto de una publicación anterior.