

Minería de datos aplicada al ingreso laboral en Ecuador: comparación de modelos MPL, Logit y Probit.

Data mining applied to labor income in Ecuador: comparison of LPM, Logit, and Probit models.

Juna-Pozo, Luis Patricio¹.

¹ Universidad Central del Ecuador; Ecuador, Quito; <https://orcid.org/0009-0005-9006-0502>; lpjuna@uce.edu.ec

¹ Autor Correspondencia

 <https://doi.org/10.63618/omd/isj/v4/n3/384>

Cita: Juna-Pozo, L. P. (2026). Minería de datos aplicada al ingreso laboral en Ecuador: comparación de modelos MPL, Logit y Probit. *Innova Science Journal*, 4(3), 814-832. <https://doi.org/10.63618/omd/isj/v4/n3/384>

Recibido: 17/02/2026

Aceptado: 13/07/2026

Publicado: 31/07/2026



Copyright: © 2026 por los autores. Este artículo es un artículo de acceso abierto distribuido bajo los términos y condiciones de la Licencia Creative Commons, Atribución-NoComercial 4.0 Internacional. (CC BY-NC).

(<https://creativecommons.org/licenses/by-nc/4.0/>)

Resumen: Este artículo analiza los determinantes del ingreso laboral en Ecuador a partir de microdatos de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU 2023), utilizando un enfoque cuantitativo basado en modelos de variable dependiente binaria. Se estiman y comparan el Modelo de Probabilidad Lineal (MPL), Logit y Probit con el objetivo de identificar los factores que influyen en la probabilidad de percibir ingresos inferiores al salario básico unificado de 450 dólares. Las variables consideradas incluyen características sociodemográficas, laborales y educativas. Los resultados muestran que la no afiliación al seguro social, la jornada laboral reducida y el sexo femenino incrementan significativamente la probabilidad de ingresos bajos, evidenciando condiciones estructurales de informalidad y desigualdad de género. En contraste, la alfabetización, el empleo en el sector privado, la residencia urbana y la jefatura del hogar actúan como factores protectores. El modelo Logit presentó el mejor ajuste estadístico, mientras que el Probit mostró mayor sensibilidad en rangos intermedios de probabilidad. La consistencia de los resultados entre modelos refuerza la robustez del análisis y aporta evidencia empírica relevante para el diseño de políticas públicas orientadas a la formalización laboral y la reducción de desigualdades salariales en el contexto ecuatoriano.

Palabras clave: Minería de datos; Modelos binarios; Ingreso laboral; Informalidad estructural; Desigualdad de género.

Abstract: This article analyzes the determinants of labor income in Ecuador using microdata from the 2023 National Employment, Unemployment, and Underemployment Survey (ENEMDU) through a quantitative approach based on binary dependent variable models. The Linear Probability Model (LPM), Logit, and Probit are estimated and compared to identify factors influencing the likelihood of earning below the unified basic wage of USD 450. The analysis incorporates sociodemographic, labor, and educational variables. The results indicate that lack of social security affiliation, shorter working hours, and being female significantly increase the probability of low income, reflecting structural conditions of informality and gender inequality. Conversely, literacy, private-sector employment, urban residence, and household headship reduce this probability and act as protective factors. The Logit model exhibited the best overall statistical fit, while the Probit model showed greater sensitivity in intermediate probability ranges. The consistency of results across models strengthens the robustness of the findings and provides empirical evidence to support public policies aimed at labor formalization and the reduction of wage inequalities in the Ecuadorian labor market.

Keywords: Data mining; Binary models; Labor income; Structural informality; Gender inequality.

1. Introducción

El estudio de los determinantes salariales en América Latina, según la Comisión Económica para América Latina y el Caribe (CEPAL, 2021), ha adquirido relevancia en el análisis económico y social debido a la persistencia de la informalidad, la desigualdad estructural y la segmentación del mercado laboral. En Ecuador, estas condiciones se expresan en brechas de ingreso asociadas a factores como el género, la afiliación al seguro social, la zona de residencia y el tipo de ocupación, tal como señalan investigaciones recientes sobre el mercado laboral ecuatoriano (López & Sarmiento, 2019). Comprender la forma en que estas variables influyen en los ingresos laborales resulta esencial para orientar políticas públicas que promuevan la equidad y la inclusión económica.

La Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) del año 2023 que constituye una fuente oficial de microdatos que permite analizar esta problemática desde un enfoque cuantitativo riguroso (INEC, 2023), de la cual ya no es susceptible a múltiples cambios o correcciones que alteren los resultados posteriores y representa un punto de partida para posteriores análisis, con base en esta información, el presente estudio desarrolla un proceso ETL para estructurar un conjunto de datos consistente, sobre el cual se estiman tres modelos econométricos: el Modelo de Probabilidad Lineal (MPL), Logit y Probit.

Diversos trabajos metodológicos han demostrado que la comparación entre modelos de respuesta binaria permite evaluar la estabilidad de los coeficientes y detectar posibles sesgos en la especificación funcional, lo que fortalece la validez interna de los resultados (Hosmer, Lemeshow y Sturdivant 2013; Gujarati y Porter 2021). Asimismo, la incorporación de técnicas de ciencia de datos en el procesamiento de encuestas sociales ha contribuido a mejorar la precisión y replicabilidad de los análisis empíricos en estudios laborales (OIT 2022; CEPAL 2022).

El propósito de este estudio es identificar los factores que explican las diferencias salariales en Ecuador durante 2023, contrastando los resultados obtenidos entre los tres modelos mencionados, este análisis busca aportar evidencia empírica que contribuya al diseño de políticas laborales basadas en datos y orientadas a reducir las desigualdades estructurales del mercado de trabajo ecuatoriano.

El interés por comprender las desigualdades de ingreso se vincula con los compromisos internacionales asociados a los Objetivos de Desarrollo Sostenible (ODS), especialmente el ODS 8 sobre trabajo decente y crecimiento económico, la literatura reciente destaca la importancia de analizar los ingresos desde una perspectiva integral que incorpore factores sociodemográficos, educativos y laborales que influyen en la probabilidad de acceder a un empleo formal y adecuadamente remunerado (OIT 2022; CEPAL 2022). En este marco, el uso de microdatos de encuestas de hogares constituye una herramienta fundamental para caracterizar las condiciones laborales en contextos territoriales específicos.

La comparación entre los modelos MPL, Logit y Probit aporta un enfoque analítico que permite interpretar las relaciones entre variables desde distintos marcos funcionales y probabilísticos, lo cual resulta especialmente útil cuando se trabaja con variables categóricas o dicotómicas en la variable dependiente, los modelos Logit y Probit han

sido ampliamente utilizados en economía y ciencias sociales para analizar decisiones binarias, ya que aseguran que las probabilidades estimadas se mantengan dentro de límites coherentes y permiten interpretar los efectos marginales en contextos no lineales (Hosmer, Lemeshow y Sturdivant 2013; Gujarati y Porter 2021). El aporte de este estudio radica en integrar principios de la ciencia de datos con herramientas econométricas para generar evidencia precisa y replicable que contribuya al diseño de políticas laborales diferenciadas y efectivas.

2. Materiales y Métodos

Este estudio se fundamenta en un enfoque cuantitativo de corte transversal, basado en microdatos de la ENEMDU correspondiente al año 2023, proporcionada por el Instituto Nacional de Estadística y Censos del Ecuador (INEC 2023). La base de datos incluye información sociodemográfica, ocupacional y de ingresos, recopilada mediante un diseño muestral probabilístico estratificado, representativo a nivel nacional, urbano y rural.

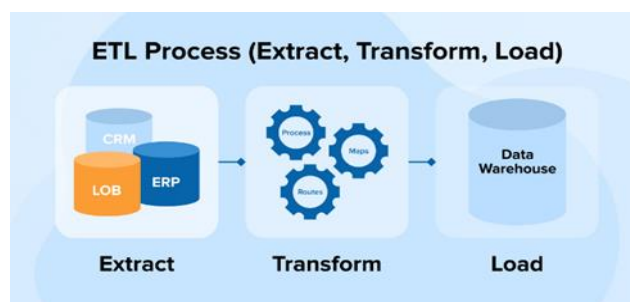
Para poder identificar las variables que contribuyan a los modelos econométricos propuestos, se realiza un proceso de análisis del estado del arte de la problemática y sus componentes teóricos, así como un análisis correlacional de las variables para observar su significancia, su asociación y su transformación para cumplir con las condiciones del modelo de variable dependiente binaria (MVDB). Estudios recientes destacan la importancia de la preparación rigurosa de datos y la validación estadística en investigaciones basadas en encuestas, especialmente cuando se aplican modelos de clasificación binaria (Silva Filho et al. 2023).

2.1. Proceso ETL

Se aplicó un proceso ETL (Extracción, Transformación y Carga) en el software estadístico SPSS V26 (Figura 1). La extracción se realizó desde archivos oficiales del INEC. La transformación incluyó limpieza de datos, recodificación de variables categóricas como sexo, afiliación al IESS y zona de residencia, además de la creación de variables dummy como sedentarismo. La carga consistió en estructurar un dataset analítico para su posterior modelado estadístico con tres modelos econométricos: MPL, Logit y Probit. La literatura reciente subraya que los procesos ETL y la limpieza de datos son determinantes para garantizar la calidad analítica y la validez de los modelos estadísticos aplicados (Sajad Hussain 2021; Martínez Fernández 2014).

Figura 1.

Proceso ETL (Extract, Transform, Load)



Nota: Elaborado por CData Software (2023).

El enfoque multivariado se eligió debido a la naturaleza exploratoria del fenómeno, permitiendo reducir la dimensionalidad de las variables mediante análisis factorial e identificar predictores significativos mediante modelos logísticos. Las técnicas cuantitativas multivariadas permiten estructurar fenómenos sociales complejos bajo principios de objetividad, replicabilidad y análisis simultáneo de múltiples variables (Hair et al., 2019).

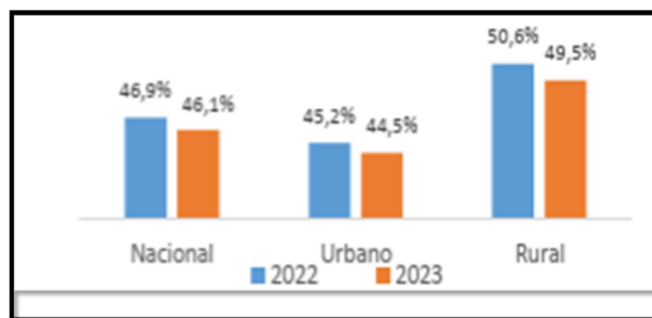
2.1.1. Extracción

En la fase de extracción, se obtuvo el archivo en formato .sav, lo que permitió preservar la estructura original de los datos y asegurar su integridad desde una fuente oficial, en trabajos recientes recomiendan que la precisión en la extracción desde fuentes homogéneas es fundamental para evitar pérdidas de información, errores de codificación y alteraciones en la estructura de los datos, especialmente en estudios basados en encuestas sociales (Alleva et al. 2021).

La ENEMDU 2023 fue seleccionada por responder a los ajustes metodológicos incorporados en dicho periodo, los cuales constituyen mejoras sustanciales respecto a versiones anteriores de la encuesta, se prevé que estos cambios se mantendrán en las futuras ediciones, permitiendo una mayor estabilidad y comparabilidad en el análisis longitudinal de indicadores laborales, de acuerdo con lo señalado en el Boletín N.º 05-2024, la tasa de participación bruta registrada en 2023 alcanzó el 46,1 % a nivel nacional, mientras que en el área urbana fue del 44,5 % y en el área rural del 49,5 % (INEC, 2024)(Figura 2.).

Figura 2.

Tasa de participación bruta a nivel nacional, urbano y rural, años 2022 y 2023



Nota: Boletín Técnico N° 05-2024-ENEMDU.

En cuanto a la metodología establecida en el Boletín Técnico Nro. 05-2024 de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU), la muestra trimestral correspondiente al año 2023 estuvo compuesta por 27 048 viviendas, lo que garantiza una representatividad estadística adecuada a nivel nacional, tanto en áreas urbanas como rurales, adicionalmente, se incluyó representatividad específica para cinco ciudades que se consideran auto-representadas dentro del diseño muestral: Quito (capital del Ecuador), Guayaquil, Cuenca, Machala y Ambato (INEC, 2024).

2.1.2. Transformación

En la fase de transformación es importante, para Juna-Luis (2025) garantiza la calidad de la información la limpieza de los datos, en este trabajo se realizó mediante el software

SPSS V26, este proceso consistió en la codificación de variables categóricas, la imputación de valores perdidos, la normalización de etiquetas, la creación de variables dummy para hacerlas comparativas y la aplicación de reglas semánticas para asegurar la coherencia analítica, la transformación permitió adaptar los datos a los requerimientos de los modelos MPL, Logit y Probit, facilitando su explotación estadística. Estudios recientes destacan que la estandarización semántica y la validación estructural son esenciales para garantizar coherencia y minimizar errores en datos sociales, especialmente en encuestas de gran escala (Maslovskaya et al. 2022; Leeuw, D. y Wooldridge, 2020).

Para realizar los cambios de variables cuantitativas se consideró el salario básico unificado de 450 dólares establecido para el año 2023, comparando este valor con los ingresos superiores a dicho umbral (Ministerio del Trabajo, 2022), para identificar posibles hábitos asociados a una vida sedentaria, se tomó como referencia la recomendación de la OMS, que señala que “los adultos deben realizar al menos 150 minutos de actividad física aeróbica de intensidad moderada a la semana, o al menos 75 minutos de actividad aeróbica de intensidad vigorosa, o una combinación equivalente de ambas” (OMS, 2020), esta directriz permitió definir la variable D_AF relacionada con la actividad física moderada.

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.

2.1.3. Carga

En la fase de carga, los datos previamente depurados fueron incorporados al entorno analítico de SPSS V26, donde se ejecutaron el análisis descriptivo de las principales variables, las correlaciones entre estas y la estimación de los modelos econométricos, esta etapa permitió asegurar la persistencia estructurada de la información, así como su trazabilidad y reutilización en análisis comparativos y predictivos, aspectos fundamentales para la replicabilidad y consistencia metodológica del estudio (Weidmann, 2019).

2.2. 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
D_Ingresos	Dicotómica (Dependiente)	Ingreso laboral mensual comparativo de \$450	1 = Menor a \$450; 0 = Otros ingresos
Edad	Númerica discreta	Edad del encuestado en años	Valor numérico
Asalariados	Númerica continua	Salario por un trabajo fijo	Valor numérico
Sexo	Dicotómica	Sexo del encuestado	1 = Mujer; 0 = Hombre
Estado_civil	Dicotómica	Estado civil del encuestado	1 = Casado; 0 = Otros
Jefe_hogar	Dicotómica	Si el encuestado es jefe de hogar	1 = Jefe de hogar; 0 = Otros
Ocupación	Dicotómica	Tipo de ocupación	1 = Empleado privado; 0 = Otros
Zona	Dicotómica	Zona geográfica de residencia	1 = Urbana; 0 = Rural
D_AF	Dicotómica	Si realiza actividad física moderada	1 = Sí; 0 = No
Instrucción	Dicotómica	Nivel educativo alcanzado	1 = Secundaria; 0 = Otros niveles

Trabajo	Dicotómica	Jornada laboral semanal	1 = menor o igual a 40 horas; 0 = mayor de 40 horas
Seguro	Dicotómica	Afiliación algún tipo de seguro social	1 = Afiliado; 0 = No afiliado
Alfabetizados	Dicotómica	Capacidad de lectura y escritura del encuestado	1 = Sí; 0 = No

Nota: Elaborado por el autor en base a los datos de la ENEMDU 2023.

2.3. Modelos econométricos

La modelización estadística constituye una herramienta fundamental para el análisis de fenómenos socioeconómicos cuando la variable de interés adopta una estructura dicotómica, como ocurre en los estudios de ingresos laborales, la aplicación de modelos econométricos para variables dependientes binarias permite estimar probabilidades condicionadas y evaluar la influencia relativa de factores individuales, laborales y contextuales, aportando evidencia empírica sólida para la interpretación de desigualdades económicas. Investigaciones recientes señalan que la comparación entre modelos lineales y no lineales mejora la robustez analítica y favorece una lectura más precisa de los determinantes asociados a la vulnerabilidad salarial en estudios basados en microdatos de encuestas (Cameron y Trivedi 2021; Long y Freese 2022). Se estimaron tres modelos econométricos para el análisis comparativo.

2.3.1. Modelo de Probabilidad Lineal (MPL)

El Modelo de Probabilidad Lineal (MPL) es una forma simplificada de modelar variables dependientes binarias, donde la respuesta toma valores de “0” y “1”, a pesar de sus limitaciones teóricas, su simplicidad lo hace útil como punto de partida en estudios empíricos, particularmente cuando se busca una primera aproximación a las relaciones entre variables explicativas y una respuesta dicotómica (Wooldridge, 2020).

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

Donde:

- Y_i es la variable dependiente binaria
- β_0 es la constante o intercepto
- β_j son los coeficientes a estimar
- ε_1 es el término de error.

La interpretación de los coeficientes es directa: un incremento unitario en X_j cambia la probabilidad de que $Y=1$ en β_j unidades, manteniendo constante el resto de variables (principio *certeris paribus*).

2.3.2. Modelo Logit

El modelo Logit resulta apropiado cuando la variable dependiente es dicotómica, ya que permite modelar la probabilidad de ocurrencia de un evento en función de múltiples variables explicativas bajo un marco probabilístico no lineal. Este enfoque garantiza que las probabilidades estimadas se mantengan dentro del intervalo lógico entre cero y uno, y facilita la interpretación de los coeficientes a través de razones de probabilidades, lo que aporta claridad analítica en estudios socioeconómicos. Investigaciones metodológicas recientes destacan que el modelo Logit ofrece estabilidad en la estimación y una interpretación consistente de los efectos marginales, especialmente en análisis basados en microdatos de encuestas laborales y sociales (Greene, 2018). El uso de modelos logísticos binarios ha demostrado ser una herramienta eficaz para analizar decisiones y comportamientos representados mediante variables dicotómicas.

en diferentes ámbitos de las ciencias sociales y económicas, en un estudio reciente desarrollado con estudiantes universitarios ecuatorianos, la regresión logística permitió identificar factores asociados a la intención emprendedora, evidenciando la utilidad de este enfoque para estimar probabilidades y evaluar la influencia de variables explicativas sobre eventos binarios (Valencia-Arias et al., 2026).

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}} \quad (2)$$

- p es la probabilidad que ocurra el evento.
- son los odds (razón de éxito vs. fracaso).
- $\ln$ logaritmo natural.
- β_0 es la constante o intercepto
- $\beta_1, \beta_2, \dots, \beta_k$ son los coeficientes de $X_1, X_2, \dots, X_k$ que son las variables independientes.

Este modelo asume una distribución logística acumulada para la variable latente. Se estimó mediante máxima verosimilitud.

2.3.3. Modelo Probit

El modelo Probit es ampliamente utilizado en estudios socioeconómicos para el análisis de variables dependientes binarias, como la probabilidad de estar empleado, percibir ingresos o acceder a servicios sociales. A diferencia del Modelo de Probabilidad Lineal, el enfoque Probit asume una distribución normal de la variable latente, lo que permite estimaciones probabilísticas acotadas entre cero y uno y una interpretación más precisa de los efectos marginales en contextos caracterizados por informalidad y segmentación estructural del mercado laboral. La literatura metodológica reciente destaca su utilidad para modelar decisiones binarias cuando se trabaja con microdatos de encuestas sociales y económicas (Perez-Truglia, 2020; Agresti, 2019). (Perez-Truglia, 2020).

$$P(Y = 1|X) = \Phi(X^T \beta) \quad (3)$$

Donde:

- $P(Y=1 | X)$: es la probabilidad de que el evento ocurra (por ejemplo, que una persona tenga ingresos).
- 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 & Porter, 2010).

2.4. Validación y pruebas estadísticas

La validación y las pruebas estadísticas permiten comprobar la consistencia y confiabilidad de los modelos econométricos estimados, estas pruebas se utilizan para verificar la asociación entre las variables explicativas y descartar problemas de multicolinealidad, garantizando la validez interna y la estabilidad de los resultados obtenidos.

2.4.1. Correlación de variables

Para la identificación de variables potencialmente significativas se consideró, en primer lugar, el estado del arte del problema en estudio. Posteriormente, para evaluar la asociación estadística y la independencia entre variables categóricas, se aplicó la prueba de Chi cuadrado de Pearson. Esta técnica no paramétrica resulta adecuada para el análisis de datos provenientes de encuestas, ya que no requiere supuestos de normalidad y permite identificar relaciones significativas entre variables cualitativas en estudios socioeconómicos basados en grandes muestras (Agresti, 2019;Field, 2017).

Tabla 2.

Variables usadas en los modelos econométricos

Variables	Estadístico	Ingresos
Edad	Correlación de Pearson	-,214**
	Sig. (bilateral)	0,000
Ingresos fijos	Correlación de Pearson	-,371**
	Sig. (bilateral)	0,000
Mujeres	Correlación de Pearson	,193**
	Sig. (bilateral)	0,000
Casados	Correlación de Pearson	-,145**
	Sig. (bilateral)	0,000
Jefe del hogar	Correlación de Pearson	-,270**
	Sig. (bilateral)	0,000
Empleados Privados	Correlación de Pearson	-,347**
	Sig. (bilateral)	0,000
Urbana	Correlación de Pearson	-,164**
	Sig. (bilateral)	0,000
Actividad física moderada	Correlación de Pearson	,070**
	Sig. (bilateral)	0,000
Secundaria	Correlación de Pearson	,090**
	Sig. (bilateral)	0,000
Trabajó al menos de 40 horas	Correlación de Pearson	,095**
	Sig. (bilateral)	0,000
Los que no aportan algún seguro	Correlación de Pearson	,608**
	Sig. (bilateral)	0,000
Saben leer o escribir	Correlación de Pearson	-,064**
	Sig. (bilateral)	0,000

Nota: Elaborado por el autor

Para el análisis se debe comprobar las siguientes hipótesis:

- Ho: No existe correlación lineal significativa entre las variables.
- Ha: Existe correlación lineal significativa entre las variables.

Siendo el valor crítico:

Si $p\text{-valor} \leq 0.05$; Rechazo Ho

Decisión: Existe correlación lineal significativa entre las variables.

Los resultados obtenidos en la Tabla 2, en este estudio revelan asociaciones estadísticamente significativas entre todas las variables, la no afiliación al seguro social presenta la correlación más fuerte ($r = .608$; $p\text{-valor} < .001$), evidenciando una relación estructural entre la informalidad laboral y la baja percepción salarial. Asimismo, variables como ingresos fijos ($r = -.371$), ocupación en el sector privado ($r = -.347$) y condición de jefe del hogar ($r = -.270$) muestran correlaciones negativas relevantes, lo cual sugiere que estas características no garantizan necesariamente ingresos superiores.

Por otro lado, el sexo femenino ($r = .193$), estado civil casado ($r = -.145$) y residencia urbana ($r = -.164$) también se vinculan significativamente con el nivel de ingresos, aunque con menor intensidad. Todas las correlaciones fueron significativas al nivel del 0.001, lo que valida su inclusión como predictores en los modelos econométricos MPL, Logit y Probit.

2.4.2. Prueba de no multicolinealidad

El análisis de multicolinealidad resulta fundamental para garantizar la validez interna de los modelos econométricos propuestos (MPL, Logit y Probit). La multicolinealidad se refiere a la presencia de correlaciones elevadas entre dos o más variables independientes, lo cual puede distorsionar la estimación de los coeficientes, aumentar los errores estándar e incluso revertir la significancia estadística de predictores relevantes (Hair et al. 2019). Para abordar este problema, se aplicó el Factor de Inflación de la Varianza (VIF, por sus siglas en inglés) sobre las variables numéricas y dicotómicas incluidas en el modelo. Valores de VIF inferiores a 5 se consideraron indicativos de no colinealidad o a su vez en el coeficiente de tolerancia, si este es cercano a 1, es un indicativo de no colinealidad (Wooldridge, 2020).

Los valores obtenidos de tolerancia y VIF (Tabla 3), indican que no existe multicolinealidad severa entre las variables independientes del modelo propuesto. Todas las tolerancias se mantienen por encima de 0.60 y los VIF por debajo de 2, lo cual está dentro de los rangos aceptables recomendados en la literatura metodológica ($VIF < 5$; tolerancia ≈ 1). La variable con menor tolerancia es "los que no aportan algún seguro" (0.680; $VIF = 1.470$), y aunque su valor es el más cercano al umbral crítico, no representa riesgo técnico. En conjunto, estos resultados confirman que las variables seleccionadas para modelar los ingresos mensuales en Ecuador aportan información independiente y relevante, fortaleciendo la robustez del enfoque econométrico planteado.

Tabla 3.

Prueba de no multicolinealidad VIF

Variable	Tolerancia	VIF
Edad	0,606	1,649
Ingresos fijos	0,840	1,190
Mujeres	0,835	1,198
Casados	0,801	1,248
Jefe del hogar	0,717	1,394
Empleados Privados	0,775	1,291
Urbana	0,921	1,085

Actividad física moderada	0,902	1,108
Secundaria	0,876	1,141
Trabajó al menos de 40 horas	0,996	1,004
Los que no aportan algún seguro	0,680	1,470
Saben leer o escribir	0,949	1,054

Nota: Elaborado por el autor

3. Resultados

Los resultados obtenidos en el presente estudio fueron procesados incorporando el factor de expansión muestral, lo que permitió ajustar los valores estimados a la representatividad poblacional definida por el diseño complejo de la ENEMDU 2023. Esta ponderación garantiza que los estadísticos descriptivos y los modelos econométricos reflejen con mayor precisión la estructura sociodemográfica del país, fortaleciendo la validez externa de los hallazgos y su aplicabilidad en el análisis de política pública, aplicados tanto en la estadística descriptiva como en los modelos econométricos a continuación:

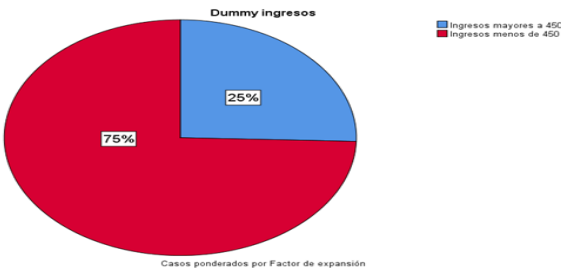
3.1. Estadística descriptiva

La estadística descriptiva resulta fundamental para comprender el comportamiento de las variables principales del estudio, ya que permite identificar patrones, tendencias y posibles sesgos en los datos analizados. En esta investigación, su aplicación facilitó la caracterización de la variable dependiente asociada al ingreso básico, así como de variables clave relacionadas con la condición laboral y la afiliación a la seguridad social, las cuales presentan asociaciones relevantes con el nivel de ingresos. La literatura metodológica reciente destaca que la estadística descriptiva constituye una etapa indispensable para resumir y organizar la información de manera clara y sistemática, garantizando la coherencia analítica previa a la estimación de modelos econométricos (Triola, 2022).

En la Figura 3, se observa que el 75% de los encuestados se encuentra en el grupo de generar ingresos menores a \$450, lo que refleja una alta concentración de población en situación económica vulnerable, 3 de cada 4 ecuatorianos tiene ingresos menores que el sueldo básico, mientras que el otros 25% superó ese lumbral de ingresos.

Figura 3.

Gráfico circular sobre ingresos

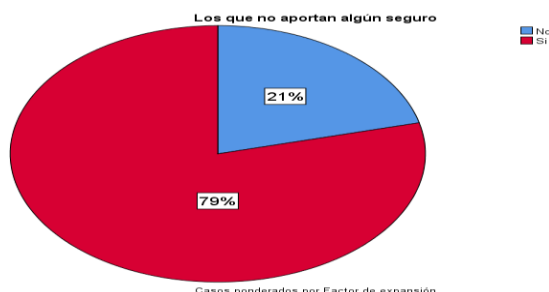


Nota: Elaborado por el autor

En la Figura 4, se observa que el 79% de los encuestados no aporta a ningún tipo de seguro social, mientras que solo el 21% sí lo hace, casi 8 de cada 10 ecuatorianos se supone que no tienen un empleo formal.

Figura 4.

Gráfico circular sobre las personas que aportan al seguro social



Nota: Elaborado por el autor

En la tabla 4, los asalariados presentan una fuerte concentración en torno al salario básico de \$450 mensuales, como lo refleja la coincidencia entre media, mediana y moda. Esta homogeneidad sugiere la presencia de un salario base estructurado en el mercado laboral, sin embargo, la dispersión salarial evidenciada por la desviación estándar de \$198,83 y la varianza superior a \$39 000, indica que existen grupos con ingresos significativamente distintos.

Tabla 4.

Medidas de tendencia centra y de dispersión del ingreso laboral

Estadísticos	
Media	450,36
Mediana	450,00
Moda	450
Desv.	198,827
Desviación	
Varianza	39532,133

Nota: Elaborado por el autor

3.2. Modelo de Probabilidad Lineal

En la Tabla 5, se observa que la variable asociada a la no afiliación a un seguro social presenta el mayor coeficiente estandarizado ($\beta = 0,429$), lo que indica una fuerte asociación positiva con la condición de ingresos cercanos o equivalentes al salario básico, consolidándose como un indicador relevante de informalidad laboral, de igual forma, pertenecer al sexo femenino ($\beta = 0,101$) y haber trabajado al menos 40 horas semanales ($\beta = 0,068$) muestran asociaciones positivas con la variable dependiente, aunque con menor magnitud relativa.

Las variables como empleo en el sector privado ($\beta = -0,134$), condición de jefe del hogar ($\beta = -0,106$) e ingresos fijos ($\beta = -0,199$) presentan coeficientes negativos, lo que evidencia una asociación inversa con la condición de ingresos bajos y sugiere una mayor vinculación con niveles de ingreso superiores al salario básico, todos los

coeficientes resultaron estadísticamente significativos (p -valor $< 0,001$), lo que respalda la capacidad explicativa del modelo MPL dentro del estudio.

Tabla 5.

Modelo de Probabilidad Lineal

Variables	Coeficientes no estandarizados		Coeficientes estandarizados		
	B	Desv. Error	Beta	t	Sig.
(Constante)	0,418	0,001		287,438	0,000
Edad	-0,002	0,0000	-	-181,103	0,000
Ingresos fijos	0,00	0,0000	-	-813,624	0,000
Mujeres	0,088	0,0000	0,101	412,316	0,000
Casados	-0,026	0,0000	-	-104,405	0,000
Jefe del hogar	-0,099	0,0000	-	-400,683	0,000
Empleados Privados	-0,143	0,0000	-	-526,201	0,000
Urbana	-0,069	0,0000	-	-311,77	0,000
Actividad física moderada	0,022	0,0000	0,02	85,793	0,000
Secundaria	0,015	0,0000	0,016	66,252	0,000
Trabajó al menos de 40 horas	0,312	0,0010	0,068	302,929	0,000
Los que no aportan algún seguro	0,458	0,0000	0,429	1578,472	0,000
Saben leer y escribir	-0,061	0,001	-	-98,918	0,000
			0,023		

Nota: Elaborado por el autor

3.3. Modelo Logit

En la Tabla 6, los resultados del modelo Logit muestran que la no afiliación a un seguro social ($\text{Exp}(B) = 3,978$), haber trabajado al menos 40 horas semanales ($\text{Exp}(B) = 6,329$) y pertenecer al sexo femenino ($\text{Exp}(B) = 1,958$) incrementan de manera significativa la probabilidad de encontrarse en la condición de ingresos bajos. Estas variables se asocian con características del mercado laboral vinculadas a informalidad, intensidad de la jornada laboral y desigualdades de género.

Las variables como empleo en el sector privado ($\text{Exp}(B) = 0,243$), condición de jefe del hogar ($\text{Exp}(B) = 0,391$), residencia en zona urbana ($\text{Exp}(B) = 0,702$) y alfabetización ($\text{Exp}(B) = 0,539$) presentan razones de probabilidades inferiores a la unidad, lo que indica una menor probabilidad de ubicarse en el grupo de ingresos bajos. La mayoría de los coeficientes resultaron estadísticamente significativos al nivel del 1 %, mientras que la variable estado civil casado fue significativa al 5 %, lo que respalda la consistencia estadística del modelo Logit para el análisis del ingreso laboral.

Tabla 6.

Modelo Logit

Variables	B	Error estándar	Wald	Sig.	Exp(B)
Edad	-0,026	0,000	41398,871	0,000	0,974
Ingresos Laborales	-0,020	0,000	802818,680	0,000	0,980
Mujeres	0,672	0,003	55882,098	0,000	1,958
Casados	-0,007	0,003	4,147	0,042	0,993
Jefe del hogar	-0,940	0,003	91568,898	0,000	0,391
Empleados Privados	-1,414	0,003	204566,469	0,000	0,243
Urbana	-0,353	0,003	12072,501	0,000	0,702
Actividad física moderada	0,492	0,003	22467,485	0,000	1,635
Secundaria	0,157	0,003	2913,558	0,000	1,170
Trabajó al menos de 40 horas	1,845	0,011	26909,682	0,000	6,329
Los que no aportan algún seguro	1,381	0,003	177585,269	0,000	3,978
Saben leer o escribir	-0,618	0,010	3905,347	0,000	0,539
Constante	9,404	0,021	195577,697	0,000	12139,664

Nota: Elaborado por el autor

3.4. Modelo Probit

En la Tabla 7, los resultados del modelo Probit indican que la no afiliación a un seguro social ($\text{Exp}(B) = 9,267$) y haber trabajado al menos 40 horas semanales ($\text{Exp}(B) = 7,545$) se asocian de manera significativa con una mayor probabilidad de ubicarse en el grupo de ingresos bajos, evidenciando vínculos con condiciones de informalidad y exigencias laborales intensivas. Asimismo, pertenecer al sexo femenino ($\text{Exp}(B) = 2,303$) incrementa significativamente la probabilidad de encontrarse en dicha condición.

Tabla 7.

Modelo Probit

Variables	B	Error estándar	Wald	Sig.	Exp(B)
Edad	-0,022	0,000	49309,039	0,000	0,978
Ingresos fijos	-0,011	0,000	462473,466	0,000	0,989
Mujeres	0,834	0,002	136916,227	0,000	2,303
Casados	-0,191	0,002	5905,762	0,000	0,826

Jefe del hogar	-0,898	0,002	139053,84 2	0,00 0	0,407
Empleado s Privados	-1,450	0,003	293462,95 2	0,00 0	0,235
Urbana	-0,739	0,003	84291,228	0,00 0	0,478
Actividad física moderada	0,297	0,003	13587,989	0,00 0	1,346
Secundaria	0,122	0,002	2887,429	0,00 0	1,129
Trabajó al menos de 40 horas	2,021	0,008	59213,305	0,00 0	7,545
Los que no aportan algún seguro	2,226	0,003	784137,30 8	0,00 0	9,267
Saben leer o escribir	-0,722	0,008	8402,660	0,00 0	0,486
Constante	4,630	0,016	86449,359	0,00 0	102,50 4

Nota: Elaborado por el autor

3.5. Comparación de los modelos econométricos

En la Tabla 8, los tres modelos econométricos presentan consistencia en la dirección de los efectos estimados para todas las variables analizadas, lo que refuerza la estabilidad de los resultados obtenidos, las variables asociadas a la informalidad laboral y a condiciones estructurales del empleo, como la no afiliación al seguro social, el sexo femenino y la intensidad de la jornada laboral, mantienen coeficientes positivos en los tres enfoques. En cambio, variables vinculadas a una mayor estabilidad económica, como el empleo en el sector privado, la condición de jefe del hogar, la residencia en zona urbana y la alfabetización, muestran efectos negativos de forma sistemática.

En términos de magnitud, el MPL presenta coeficientes relativamente menores, como se observa en la variable no aportar al seguro social ($B = 0,458$) y en el sexo femenino ($B = 0,088$), lo que refleja las limitaciones inherentes a su estructura lineal para capturar relaciones no lineales, el modelo Logit incrementa la magnitud de estos efectos al expresarlos en una escala logarítmica de probabilidades, evidenciando coeficientes más elevados para las mismas variables ($B = 1,381$ y $B = 0,672$, respectivamente) y el modelo Probit reproduce este patrón general, con coeficientes particularmente altos en variables como no aportar al seguro social ($B = 2,226$) y haber trabajado al menos 40 horas semanales ($B = 2,021$), lo que sugiere una mayor sensibilidad en los rangos intermedios de la distribución probabilística.

Las variables empleo en el sector privado y jefatura del hogar conservan coeficientes negativos robustos y de magnitud comparable en los modelos Logit y Probit, indicando una menor probabilidad de ubicarse en el grupo de ingresos bajos cuando estas condiciones están presentes. La convergencia en los signos y la coherencia relativa de las magnitudes entre los tres modelos refuerzan la solidez empírica del análisis y justifican el uso complementario de los enfoques MPL, Logit y Probit para una comprensión integral de los determinantes del ingreso laboral.

Tabla 8.

Comparación de los modelos econométricos MPL, Logit y Probit

Variable	MPL (B)	Logit (B)	Probit (B)
Edad	-0,002	-0,026	-0,022
Ingresos fijos	0	-0,020	-0,011
Mujeres	0,088	0,672	0,834
Casados	-0,026	-0,007	-0,191
Jefe del hogar	-0,099	-0,940	-0,898
Empleados Privados	-0,143	-1,414	-1,450
Urbana	-0,069	-0,353	-0,739
Actividad física moderada	0,022	0,492	0,297
Secundaria	0,015	0,157	0,122
Trabajó menos de 40 horas	0,312	1,845	2,021
No aporta al seguro social	0,458	1,381	2,226
Alfabetización (sabe leer/escribir)	-0,061	-0,618	-0,722

Nota: Elaborado por el autor

En la Tabla 9, se evidencian diferencias claras en la capacidad explicativa de los modelos estimados, el modelo Logit presenta el mayor nivel de ajuste, reflejado en un R^2 de Nagelkerke de 0,776, seguido por el modelo Probit con un valor de 0,610, mientras que el Modelo de Probabilidad Lineal muestra un desempeño inferior, con un R^2 de Cox y Snell de 0,467; estos resultados indican una mayor capacidad del modelo Logit para capturar la variabilidad asociada a la condición de ingresos bajos.

Los valores del logaritmo de la verosimilitud confirman diferencias en la adecuación relativa de los modelos, siendo el Logit el que muestra una mejor aproximación comparativa a los datos observados, la prueba de Hosmer y Lemeshow resulta estadísticamente significativa en los modelos Logit y Probit (p -valor $< 0,001$), lo que sugiere discrepancias entre valores observados y esperados, un comportamiento común en muestras de gran tamaño, la evidencia empírica respalda el uso del modelo Logit como referencia principal, mientras que los modelos Probit y MPL aportan consistencia comparativa y robustez al análisis.

Tabla 9.

Estadísticos de representatividad de los mode

Estadístico	Logaritmo de la veros. -2	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke	Prueba de Hosmer y Lemeshow	
				Chi-cuadra	Sig.
MPL		0,467			0,000
Logit	4108128,7	0,527	0,776	49674551,3	0,000

Probit	6374075,53	0,414	0,61	150810,728	0,000
--------	------------	-------	------	------------	-------

Nota: Elaborado por el autor

4. Discusión

Los resultados obtenidos en los tres modelos econométricos (MPL, Logit y Probit) coinciden en señalar que la falta de aporte al seguro social es el principal determinante de ingresos bajos en Ecuador. Este hallazgo refuerza la literatura reciente sobre el empleo informal como fenómeno estructural que limita el acceso a derechos sociales y estabilidad económica (Espejo, 2022). La magnitud del coeficiente en el modelo Probit ($B = 2,226$) sugiere que la informalidad no solo es una condición laboral, sino un factor de exclusión distributiva, esta evidencia respalda la necesidad de políticas públicas orientadas a la formalización progresiva del empleo, especialmente en sectores vulnerables.

La variable “sexo”, analizada con referencia a las mujeres, presenta un coeficiente positivo y estadísticamente significativo en todos los modelos estimados, lo que evidencia una mayor probabilidad de pertenecer al grupo de ingresos bajos. Este resultado se articula con el efecto de trabajar menos de 40 horas semanales, reforzando la precariedad laboral como un eje persistente de desigualdad estructural en el mercado de trabajo. Estudios recientes sobre Ecuador muestran que, aun controlando por educación, experiencia laboral y características ocupacionales, persiste una brecha salarial de género atribuible a factores no observables, vinculados a discriminación salarial y segmentación ocupacional. En este sentido Linthon-Delgado y Méndez-Heras (2022), confirman que una parte sustancial de la brecha en el país no se explica por el capital humano, sugiriendo la persistencia de prejuicios de género que limitan el acceso de las mujeres a empleos de mayor calidad y estabilidad. Tales hallazgos, reflejados en los coeficientes elevados del modelo Probit, respaldan la necesidad de políticas públicas orientadas a la equidad laboral, la formalización del empleo y la corresponsabilidad social.

En variables como “alfabetización”, “zona urbana”, “empleo privado” y “ser jefe del hogar” se presentan coeficientes negativos significativos en los tres modelos, lo que indica su rol protector frente al rezago salarial. Este hallazgo es consistente con la tesis de la Comisión Económica para América Latina y el Caribe (CEPAL, 2022), la cual sostiene que en la región persisten brechas estructurales donde el entorno urbano y el nivel de instrucción actúan como determinantes críticos de la productividad y el ingreso. El entorno urbano facilita el acceso a servicios, redes laborales y oportunidades educativas, mientras que el capital humano, medido por la alfabetización, funciona como un amortiguador frente a la vulnerabilidad económica. Estos resultados refuerzan la importancia de políticas territoriales que fortalezcan las capacidades individuales y colectivas para reducir las desigualdades que caracterizan al mercado laboral latinoamericano (CEPAL, 2022).

Los resultados Rubina-López et al. (2025) señalan que la digitalización y la IA transforman con rapidez los modelos pedagógicos institucionales. Este panorama difiere parcialmente de lo hallado por Juna-Pozo (2025), cuyo estudio demuestra que, a nivel ciudadano, la aceptación tecnológica no es homogénea, sino que está condicionada por el nivel educativo y factores actitudinales. Así, mientras las instituciones proyectan un futuro digitalizado e innovador, la población mantiene una marcada incertidumbre y

preocupaciones críticas respecto a la regulación, el empleo y el impacto social de estas tecnologías (Torres et al., 2024).

5. Conclusiones

En base a los resultados obtenidos de los modelos econométricos MPL, Logit y Probit aplicados a la ENEMDU 2023, se concluye que la vulnerabilidad salarial en Ecuador responde principalmente a factores estructurales como la informalidad laboral, la precariedad en las jornadas y la desigualdad de género, la variable “no aporta al seguro social” se consolidó como el predictor más robusto del rezago distributivo, al presentar los coeficientes positivos más elevados en los tres modelos ($\beta = 0,458$ en MPL; $\beta = 2,226$ en Logit; $\beta = 1,381$ en Probit), de manera consistente, ser mujer y trabajar menos de 40 horas semanales incrementan la probabilidad de pertenecer al grupo de ingresos bajos, lo que refuerza la evidencia sobre la feminización de la pobreza y la segmentación laboral en el país.

En contraste, la alfabetización, el empleo en el sector privado, la residencia urbana y la condición de jefe de hogar actúan como factores protectores, reduciendo significativamente dicha probabilidad, estos hallazgos sugieren que el capital humano y el entorno territorial son elementos clave para amortiguar las desigualdades económicas.

Desde el punto de vista metodológico, el modelo Logit presentó el mejor ajuste estadístico (Pseudo R^2 de Nagelkerke = 0,776), posicionándose como la herramienta más robusta para este análisis, el modelo Probit aportó mayor sensibilidad en rangos intermedios de riesgo, mientras que el MPL funcionó como una aproximación exploratoria inicial, la triangulación de los tres enfoques fortalece la validez interna del estudio y aporta evidencia empírica sólida.

Los resultados reflejan la necesidad de orientar las políticas públicas a la formalización del empleo, la reducción de brechas de género y el fortalecimiento del capital humano, contribuyendo al diseño de estrategias socioeconómicas más equitativas y sostenibles en el contexto ecuatoriano.

Referencias Bibliográficas

- Agresti, A. (2019). *An introduction to categorical data analysis* (3rd ed.). John Wiley & Sons. <https://www.wiley.com/en-us/An+Introduction+to+Categorical+Data+Analysis%2C+3rd+Edition-p-9781119405283>
- Allewa, G., Falorsi, P., Righi, P., & Petrarca, F. (2021). Measuring the accuracy of aggregates computed from a statistical register. *Journal of Official Statistics*, 37(2), 481–503. <https://doi.org/10.2478/jos-2021-0021>
- Almeida, F., & Duarte Santos, J. (2020). The role of artificial intelligence in human resources. *International Journal of Social Sciences and Management Research*, 7(2), 17–27. <https://doi.org/10.32861/ijssmr.72.17.27>
- Cameron, A. C., & Trivedi, P. K. (2021). *Microeconometrics using Stata: Volume I: Regression analysis*. Stata Press. <https://www.stata.com/bookstore/microeconometrics-stata/>

- Comisión Económica para América Latina y el Caribe (CEPAL). (2021). *Informalidad laboral y la matriz de la desigualdad social: Políticas para su abordaje*. https://www.cepal.org/sites/default/files/events/files/01_lais_abramo.pdf
- Comisión Económica para América Latina y el Caribe (CEPAL). (2022). *Panorama Social de América Latina y el Caribe 2022: La transformación de la educación como eje central de la inclusión social*. <https://www.cepal.org/es/publicaciones/48518-panorama-social-america-latina-caribe-2022-la-transformacion-la-educacion>
- Espejo, A. (2022). *Informalidad laboral en América Latina: Propuesta metodológica para su identificación a nivel subnacional*. Comisión Económica para América Latina y el Caribe (CEPAL). <https://www.cepal.org/es/publicaciones/47726-informalidad-laboral-america-latina-propuesta-metodologica-su-identificacion>
- Field, A. (2017). *Discovering statistics using IBM SPSS Statistics* (5th ed.). SAGE Publications.
- Greene, W. H. (2018). *Econometric analysis* (8th ed.). Pearson.
- Gujarati, D. N., & Porter, D. C. (2010). *Econometría* (5.ª ed.). McGraw-Hill Education.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley. <https://doi.org/10.1002/9781118548387>
- Instituto Nacional de Estadística y Censos (INEC). (2023). *Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) 2023*. <https://www.ecuadorencifras.gob.ec/enemdu-anual-2023/>
- Instituto Nacional de Estadística y Censos (INEC). (2024). *Boletín técnico N.º 05 ENEMDU-2024*. https://www.ecuadorencifras.gob.ec/documentos/web-inec/EMPLEO/2024/Boletines/Boletin_Tecnico_05-2024-ENEMDU.pdf
- Juna-Pozo, L. P. (2025). Percepciones sobre la inteligencia artificial analizadas con componentes principales y regresión multinomial. *Innova Science Journal*, 3(número especial 1), 289–302. <https://doi.org/10.63618/omd/isj/v3/nE1/194>
- Linthon-Delgado, D. E., & Méndez-Heras, L. B. (2022). Descomposición de la brecha salarial de género en el Ecuador. *Revista Mexicana de Economía y Finanzas*, 17(1), 1–23. <https://doi.org/10.21919/remef.v17i1.706>
- Long, J. S., & Freese, J. (2023). *Regression models for categorical dependent variables using Stata* (3rd ed.). Stata Press.
- López, J., & Sarmiento, G. (2019). Determinantes de la brecha salarial en Ecuador: Análisis bajo un modelo minceriano relacionado con variables agregadas dummy. *Espiraes Revista Multidisciplinaria de Investigación Científica*, 3(29), 70–85. <https://doi.org/10.31876/er.v3i29.591>
- Martínez-Fernández, J. R. (2014). *Introducción al análisis de datos cuantitativos en ciencias sociales*.

<https://www.researchgate.net/publication/262158510> Introduccion al Analisis de Datos Cuantitativos en Ciencias Sociales

Maslovskaya, O., Struminskaya, B., & Durrant, G. (2022). The future of online data collection in social surveys: Challenges, developments and applications. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 185(3), 768–772. <https://doi.org/10.1111/rssa.12895>

CONFLICTO DE INTERESES

“Los autores declaran no tener ningún conflicto de intereses”.