

**Determinantes de la informalidad laboral en Norte de Santander: un análisis comparativo
2021–2024 a partir de la GEIH**

Hender Javier Gómez Quintero

Trabajo de grado para optar el título de Economista

Director

Yoana Patricia Bermúdez González

Doctorado en Economía

Codirector

Leonardo Javier Caraballo

Magíster Economía, Medio Ambiente y Recursos Naturales

Universidad Santo Tomás, Bucaramanga

División en Ciencias Económicas y Administrativas

Facultad de Economía

2026

Agradecimientos

Expreso mi más profundo agradecimiento a los docentes que acompañaron y contribuyeron al desarrollo de esta investigación. En particular, a la economista Yoana Bermúdez, por su valioso apoyo académico, el suministro de literatura especializada y sus aportes conceptuales, los cuales fueron fundamentales para el fortalecimiento teórico del estudio.

Asimismo, manifiesto mi especial agradecimiento al economista Leonardo Caraballo y al docente Henry Sebastián Rangel por su orientación permanente, su disposición para compartir conocimientos, sus valiosas sugerencias y las rigurosas correcciones realizadas durante todo el proceso. Su acompañamiento fue determinante para garantizar la calidad académica del trabajo y para la culminación exitosa de esta investigación.

Además, agradezco a mi familia por su respaldo constante a lo largo de mi carrera universitaria. De manera especial, agradezco a mi madre, María Quintero, a mi abuela, Amparo Borbón, y a mi padre, Hender Gómez, por brindarme su apoyo incondicional, su paciencia y por ser el pilar fundamental para alcanzar esta meta.

Contenido

Introducción	10
1. Determinantes de la informalidad laboral en Norte de Santander: un análisis comparativo 2021–2024 a partir de la GEIH.....	13
1.1 Planteamiento del problema	13
1.2 Justificación.....	14
1.3 Objetivos	15
1.3.1 Objetivo general	15
1.3.2 Objetivos específicos.....	15
2. Marco referencial	16
2.1 Marco teórico	16
2.1.1 Fundamentos de la Teoría del Mercado Laboral	18
2.1.2 Teorías de la informalidad.....	20
2.1.3 Teoría del Capital Humano	22
2.1.4 La Informalidad en el Contexto de la Migración	23
2.1.5 Contexto estructural de la informalidad laboral en Colombia.....	25
2.1.6 Contexto estructural de la informalidad laboral en Norte de Santander.....	26
2.2 Marco conceptual	28
2.3 Marco legal.....	30
3. Método.....	31
3.1 Modelo Econométrico	32
3.2 Metodología y Proceso de Análisis en Modelo.....	35
3.2.1 Variables Clave	35
3.2.2 Preparación de la Base de Datos	38

3.2.3 Construcción de variables del modelo principal.....	39
3.2.4 Especificación del Modelo de Probabilidad Logit	41
3.2.5 Estimación de Efectos Marginales Promedio (AME) – Probabilidades Predichas.....	42
3.2.6 Limitaciones y consideraciones metodológicas.	43
4. Resultados.....	45
4.1 Tasa de informalidad laboral.....	45
4.2 Resultados del modelo Logit.....	46
4.3 Efectos Marginales Promedio	47
4.3.1 Efecto del Nivel Educativo sobre la Probabilidad de Informalidad	49
4.3.2 Efecto de la Migración sobre la Probabilidad de Informalidad.....	50
4.3.3. Efecto de la Edad sobre la Probabilidad de Informalidad	52
4.3.4 Diferencias por Sexo en la Probabilidad de Informalidad.....	53
4.3.5 Efecto del Estrato Socioeconómico sobre la Probabilidad de Informalidad	55
4.3.6 Efecto de la Ocupación sobre la Probabilidad de Informalidad	56
4.3.7 Síntesis de los resultados	58
5. Prueba de Wald: validación global del modelo	59
6. Prueba de robustez.....	61
7. Informalidad laboral en el contexto postpandemia (2021–2024).....	63
8. Recomendaciones de Política Pública	65
9. Conclusiones.....	66
Referencias.....	69

Lista de tablas

Tabla 1. <i>Presentación de variables para cada año</i>	36
Tabla 2. <i>Presentación de coeficientes de variables y significancia modelo 2024.</i>	46
Tabla 3. <i>Presentación de coeficientes de variables y significancia modelo 2021.</i>	47

Lista de figuras

Figura 1. <i>Resultados AME año 2021.</i>	48
Figura 2. <i>Resultados AME año 2024.</i>	49
Figura 3. <i>Gráfico Probabilidad de informalidad según el nivel educativo.</i>	49
Figura 4. <i>Probabilidad de informalidad: Migrante VS Local</i>	52
Figura 5. <i>Probabilidad de informalidad según la edad</i>	53
Figura 6. <i>Evolución de la brecha de informalidad por la variable "Sexo"</i>	55
Figura 7. <i>Resultados de la Prueba Wald en R.</i>	60

Resumen

La dinámica fronteriza y la migración, hacen del trabajo informal uno de los principales problemas socioeconómicos del departamento de Norte de Santander. El presente estudio pretende analizar los determinantes sociodemográficos y laborales de la informalidad para los años 2021 y 2024 mediante un enfoque cuantitativo y un modelo de regresión logística con diseño muestral complejo aplicado a microdatos de la GEIH. Las variables analizadas mostraron que la probabilidad de pertenecer al sector informal tiende a disminuir si se tiene mayor nivel educativo y mayor edad y, en cambio, tiende a aumentar si se es mujer, se tiene estatus migratorio (analizado por lugar de nacimiento) y se ejercen ciertas ocupaciones. La informalidad descendió del 83,17% en 2021 al 78,88% en 2024 pero controlando por características individuales, la recuperación se muestra heterogénea y se mantienen y se profundizan las disparidades estructurales. Estos resultados demuestran la importancia y la necesidad de reforzar las políticas orientadas a la formalización del empleo y al mejoramiento de la calidad del trabajo en la región.

Palabras clave: informalidad, GEIH, Norte de Santander

Abstract

Border dynamics and migration make informal work one of the main socioeconomic problems in the department of Norte de Santander. This study aims to analyze the sociodemographic and labor determinants of informality for the years 2021 and 2024 using a quantitative approach and a logistic regression model with a complex sampling design applied to microdata from the GEIH (Integrated Household Survey). The variables analyzed showed that the probability of belonging to the informal sector tends to decrease with higher levels of education and older age, while it tends to increase if one is female, has migratory status (analyzed by place of birth), and holds certain occupations. Informality decreased from 83.17% in 2021 to 78.88% in 2024, but when controlling for individual characteristics, the recovery appears heterogeneous, and structural disparities persist and deepen. These results demonstrate the importance and necessity of strengthening policies aimed at formalizing employment and improving the quality of work in the region.

Keywords: informality, GEIH, Norte de Santander

Glosario

DANE: entidad encargada de producir y comunicar información estadística oficial para Colombia (DANE, s.f.).

Economía Informal: se define como el segmento del mercado laboral caracterizado por una baja dotación de capital, uso de tecnologías rudimentarias y una baja productividad, donde los trabajadores operan en condiciones de precariedad para garantizar su subsistencia (Portes y Haller, 2004).

GEIH (Gran Encuesta Integrada de Hogares): principal instrumento estadístico en Colombia para recolectar información sobre mercado laboral, ingresos y condiciones de vida (DANE, s.f.).

Seguridad Social: Sistema de Protección que la sociedad proporciona a sus miembros para asegurar el acceso a asistencia médica y garantizar la seguridad del ingreso en caso de vejez, enfermedad o desempleo (OIT, 2021).

Variable dependiente: es la variable objeto de estudio, cuyo comportamiento se busca explicar o predecir a través de otras variables. En modelos Logit de informalidad, suele ser una variable binaria (1 si es informal, 0 si es formal) (Gujarati y Porter, 2010).

Variable independiente: También conocida como variable explicativa o regresor; es aquel factor que influye, afecta o determina los cambios en la variable dependiente dentro de un modelo estadístico (Wooldridge, 2015).

Introducción

El empleo informal sigue siendo uno de los principales problemas de fondo que el mercado laboral colombiano no resuelve, por su alto nivel de empleo no protegido, baja productividad y falta de protección laboral. Eso queda más que claro en el departamento de Norte de Santander, donde la dinámica fronteriza, la dependencia del comercio entre los dos países y los flujos migratorios han configurado un entorno expuesto a la expansión del empleo informal. Las altas tasas de informalidad, persistentes en el país, muestran limitaciones del entorno laboral para ofrecer empleo de calidad, además, también evidencian disparidades en el acceso a mejores oportunidades laborales.

Se han llevado a cabo varios estudios sobre la informalidad desde la perspectiva estructural e institucional, que han demostrado que tiene su origen en la escasa creación de empleo formal y en los problemas de acceso a la formalización. En Colombia, la literatura ha señalado el papel del capital humano, la estructura productiva y las condiciones territoriales como determinantes de este fenómeno. Sin embargo, en las zonas de frontera como el caso de Norte de Santander, la informalidad adquiere características particulares asociadas a la migración, la volatilidad económica y la alta participación en actividades de subsistencia, lo que justifica la necesidad de estudios específicos para comprender sus determinantes en los contextos locales.

En este contexto, la siguiente pregunta de investigación se plantea: ¿Cuáles son los determinantes sociodemográficos y laborales de la informalidad en Norte de Santander y cómo cambian sus efectos sobre la probabilidad de estar en informalidad entre los años 2021 y 2024?, a partir de esta pregunta el presente estudio tiene como propósito analizar los factores que inciden en la probabilidad de informalidad laboral en el departamento, para lo cual se consideran variables como la edad, el sexo, el nivel educativo, la ocupación, el estrato socioeconómico y la migración,

esta última estimada a partir del lugar de nacimiento. El objetivo del estudio es identificar y cuantificar el efecto de las variables mencionadas sobre la inserción en el empleo informal, bajo la hipótesis de que existen diferencias significativas en la probabilidad de informalidad según las características individuales y laborales de los trabajadores.

La importancia de este estudio radica en que aporta al conocimiento de la informalidad desde una óptica empírica y regional. Desde la perspectiva académica, el estudio ofrece evidencia actualizada con microdatos y técnicas econométricas en encuestas diseñadas con muestras complejas, lo cual respalda la validez de los resultados. Los resultados, en el ámbito social y económico, permiten la identificación de grupos poblacionales con mayor condición de vulnerabilidad, siendo fundamental para el diseño de políticas públicas focalizadas orientadas a la formalización laboral y mejora de la calidad del empleo.

Para abordar el problema de investigación se utilizó un enfoque cuantitativo explicativo con base en microdatos de la Gran Encuesta Integrada de Hogares (GEIH) de los años 2021 y 2024. La metodología consiste en estimar modelos de regresión logística con diseño muestral complejo, complementados con la estimación de efectos marginales y probabilidades predichas, lo que permite interpretar mejor los resultados en términos de cambios en la probabilidad de informalidad.

Por último, el documento se estructura de la siguiente manera: en el primer apartado se expone el marco teórico, donde se abordan los principales enfoques conceptuales sobre la informalidad laboral y su contexto en Norte de Santander; luego se desarrolla la sección metodológica, en la cual se describen las fuentes de información, la construcción de variables y la especificación del modelo econométrico; en la sección de resultados se presentan los hallazgos descriptivos y econométricos, incluyendo el análisis de efectos marginales, pruebas de robustez y

validación del modelo; finalmente, se presentan las recomendaciones de política pública derivadas del estudio y las conclusiones.

1. Determinantes de la informalidad laboral en Norte de Santander: un análisis comparativo 2021–2024 a partir de la GEIH

1.1 Planteamiento del problema

El sector informal es uno de los mayores problemas del mercado laboral colombiano, con consecuencias en la productividad, la estabilidad económica y el acceso a la protección social de los trabajadores. En la actualidad, el 55,3% de los ocupados se encuentra en el sector informal, evidenciando las dificultades para generar empleo formal y mejorar la inserción laboral (Departamento Administrativo Nacional de Estadística, 2026). La situación es más grave en las ciudades que están en la frontera como la de Norte de Santander, donde la dinámica económica y las pautas migratorias han ayudado a que se genere este fenómeno.

En esta dirección la informalidad no solo es económica sino también social en relación a peores ingresos, menores niveles de estabilidad en el trabajo, menor cobertura en sistemas de seguridad social. Si bien la literatura ha aportado bastante para explicar los factores que explican la informalidad en Colombia, todavía existen vacíos en el análisis a nivel regional que integren de manera conjunta variables sociodemográficas y laborales bajo metodologías que consideren el complejo diseño muestral de encuestas como la Gran Encuesta Integrada de Hogares (GEIH). También existe poca evidencia que haya comparado la evolución reciente de estos determinantes, en especial durante el período posterior a la pandemia.

Por otra parte, el papel de la migración (aproximado por el lugar de nacimiento) junto con variables como la edad, el sexo, el nivel educativo, la ocupación y el estrato socioeconómico, no se ha analizado de forma integral en el contexto de Norte de Santander.

En este sentido, el presente estudio busca aportar mayor evidencia a través de un análisis comparativo entre los años 2021 y 2024, utilizando modelos de regresión logística con un diseño de muestra compleja, cálculo de efectos marginales y probabilidades predichas, lo que produce estimaciones más precisas y comparables y evidencia empírica relevante para la formulación de políticas públicas.

El problema de investigación consiste, por tanto, en conocer cómo los diferentes factores sociodemográficos y laborales inciden la informalidad laboral y cómo estos efectos han evolucionado en el tiempo. La siguiente pregunta de investigación es: ¿Cuáles son los determinantes sociodemográficos y laborales de la informalidad en Norte de Santander y cómo cambian sus efectos sobre la probabilidad de estar en informalidad entre los años 2021 y 2024?

1.2 Justificación

La informalidad laboral en Colombia es sin lugar a dudas uno de los problemas más graves del mercado laboral. Restringe el acceso a la protección social y daña la productividad del país. En Norte de Santander esto toma características particulares por su contexto geográfico, más los continuos flujos migratorios y una economía local muy ligada a actividades informales, lo que genera un entorno donde la precariedad laboral es muy evidente. Estos factores aumentan las brechas de bienestar, lo que hace necesario abordar de manera territorial y rigurosa el estudio del fenómeno.

Desde el ángulo teórico-metodológico la investigación toma los determinantes de la informalidad mediante el análisis de variables demográficas claves. A fin de garantizar la robustez de los resultados, se aplican modelos de regresión logística que toman en cuenta la complejidad

del diseño muestral. Este enfoque permite que las estimaciones no sean solo promedios, sino datos representativos que agregan valor científico al estudio en el análisis de encuestas de hogares.

En formación profesional, el desarrollo del presente estudio ha abierto la puerta para consolidar competencias técnicas en el manejo de microdatos y uso de la econometría aplicada. El objetivo no es sólo exponer los resultados, sino desarrollar competencias analíticas para la toma de decisiones basadas en la evidencia. Una identificación clara de los factores que perpetúan la informalidad es una herramienta certera para orientar la política pública regional hacia los grupos más vulnerables.

El interés de este trabajo es que permite comparar la evolución del mercado laboral entre los años 2021 y 2024. El contraste temporal permite separar las fluctuaciones macroeconómicas del país, por una parte, de los efectos debidos a las características individuales de los trabajadores de la región, por otra.

1.3 Objetivos

1.3.1 Objetivo general

Analizar usando datos de la GEIH, cómo la edad, el sexo, el nivel educativo, la ocupación, el estrato y la migración según el lugar de nacimiento influyen en la probabilidad de informalidad en Norte de Santander, comparando los años 2021 y 2024.

1.3.2 Objetivos específicos

Estimar modelos Logit con diseño muestral para los años 2021-2024 y calcular los efectos marginales de las variables explicativas.

Comparar la probabilidad de informalidad predicha entre 2021 y 2024 y hacer pruebas de robustez usando la variable “cotiza en salud” como definición de informalidad.

Formular recomendaciones estratégicas de política pública para el contexto de Norte de Santander, fundamentadas en los resultados del estudio, con el fin de orientar procesos de formalización laboral en la región.

2. Marco referencial

2.1 Marco teórico

Desde distintas corrientes teóricas se ha estudiado la informalidad laboral, con el fin de explicar por qué persiste en las economías en vías de desarrollo. La informalidad, señala la OIT desde un punto de vista estructuralista, se genera por la limitada capacidad de absorción de mano de obra del sector formal, dando lugar a la generación de actividades económicas de baja productividad (Tokman, 2001). Esta forma de abordar el tema es particularmente importante en contextos regionales donde la estructura productiva es débil y con escasas posibilidades de generar empleo formal. Para De Soto (1986), la informalidad es una respuesta racional a un entorno regulatorio complejo, con altos costos de formalización y barreras burocráticas. En ese sentido, la informalidad no es sólo respuesta a fallas del mercado laboral, sino también a falencias institucionales que desincentivan la formalización. Estas dos miradas no son excluyentes sino complementarias, sobre todo en territorios con características estructurales y regulatorias propias.

Para Norte de Santander estos enfoques teóricos cobran mayor relevancia, por tratarse de una zona fronteriza. Diversos estudios han señalado que las ciudades fronterizas presentan niveles de informalidad laboral persistentemente elevados, explicables por factores económicos, sociales

y geográficos específicos. Por ejemplo, Ramírez et al. (2016), encuentran que la informalidad en Norte de Santander supera el 60 % de los ocupados y se relaciona con variables como la pobreza, la desigualdad y la baja participación del sector industrial en la economía regional. Este patrón de comportamiento evidencia el predominio de actividades de poca escala, productividad y poca diversificación productiva.

En Colombia, la informalidad se ha vinculado históricamente a cambios estructurales del mercado laboral como los procesos de urbanización acelerada y la migración interna producto del conflicto armado (Ocampo, 2005). Pero en el departamento de Norte de Santander ese fenómeno adquiere características especiales, en función de su dependencia de la dinámica económica de la frontera con Venezuela. La fuerza laboral se ha visto obligada a reconfigurar sus estrategias de acceso al empleo formal influenciada por ciclos económicos en la zona producto del comercio transfronterizo, el tipo de cambio y últimamente la crisis venezolana.

En la última década se ha venido presentando un incremento de la informalidad en Norte de Santander, resultado de impactos externos producto de la crisis económica y política que enfrenta Venezuela desde el año 2015. Este contexto ha propiciado importantes flujos migratorios hacia el departamento, lo que ha generado presión sobre el mercado laboral y ha impulsado las actividades informales. Según el DANE (2024), la informalidad laboral persiste en niveles muy altos en Colombia, particularmente en las zonas de menor desarrollo económico, lo que refleja brechas estructurales en el acceso al empleo formal. En el caso particular de Norte de Santander esta situación es más grave, porque hay una alta movilidad poblacional y una limitada capacidad del sector productivo para generar empleo formal.

La informalidad es impulsada por factores como el relativo aislamiento del departamento respecto a los principales centros económicos del país, los elevados costos logísticos y la baja

articulación productiva. En este sentido, los resultados evidencian que la informalidad laboral en las regiones fronterizas es un problema estructural y persistente, derivado tanto de fallas del mercado como de debilidades institucionales (Ramírez et al., 2016).

2.1.1 Fundamentos de la Teoría del Mercado Laboral

El análisis del mercado laboral se origina en el pensamiento económico clásico, siendo Adam Smith ([1776] 2004), la principal referencia para vincular la división del trabajo con la determinación salarial. El salario, según Smith, no es una cosa que nazca al azar, sino que procede de los resultados de la negociación entre los trabajadores y los empleadores, y está influida por la escasez de mano de obra y por las necesidades de subsistencia. Fue continuada esta línea por David Ricardo con su “Ley de Hierro” que afirma que a largo plazo las remuneraciones tienden a bajar hasta el mínimo de subsistencia. Fue Alfred Marshall quien le dio rigor analítico a estas ideas a finales del siglo XIX, introduciendo el enfoque marginalista ([1890] 2013). En el modelo de equilibrio de este autor, el salario se determina por el punto de intersección entre la productividad marginal de la empresa y la valoración subjetiva del ocio por parte del individuo.

En la visión neoclásica contemporánea este análisis se consolida al considerar al factor trabajo como un mercado de factores convencional. La oferta de trabajo se concibe como una decisión individual de asignación de tiempo orientada a la maximización de la utilidad donde existe una relación directa entre el salario real y la disposición a trabajar (McConnell et al., 2017). Por otra parte, la demanda es vista como una función de la empresa, esto es, las empresas contratan mano de obra hasta que el costo de una unidad adicional de trabajo se iguala al ingreso que genera su productividad. Bajo el supuesto de competencia perfecta, este mecanismo debería purgar el mercado, asegurando el pleno empleo al salario de equilibrio.

Esta visión de mercado perfectamente competitivo, sin embargo, presenta severas limitaciones cuando se aplica a economías en desarrollo como la colombiana. De acuerdo con la teoría de la segmentación del mercado laboral de Doeringer y Piore (1971), el mercado laboral no es un todo homogéneo, sino que se fracciona en sectores que operan con 4 lógicas distintas. El sector primario o formal se caracteriza por la estabilidad laboral, altos salarios y protección legal, mientras que el sector secundario o informal se caracteriza por la inestabilidad y salarios de subsistencia. Con esta segmentación, el equilibrio salarial no es único ni eficiente, existen barreras institucionales y económicas que impiden la movilidad de los trabajadores entre sectores, generando una dualidad donde el mercado laboral no puede corregir sus desequilibrios a través del ajuste de precios y que hace de la informalidad un rasgo permanente.

Por último, las fallas del mercado laboral se concretan en el desempleo estructural y en la rigidez salarial. Este tipo de desempleo se produce, según Cahuc y otros (2014), por la desconexión entre las competencias demandadas por el aparato productivo y las poseídas por la oferta laboral. Asimismo, la existencia de salarios de eficiencia (aquellos por encima del equilibrio que se pagan para motivar la productividad) y las elevadas cargas prestacionales hacen que el costo de la formalidad sea prohibitivo. Así, por estas rigideces y limitaciones del mercado, parte de la mano de obra que no logra insertarse en el sector formal, termina yendo a la informalidad. Por consiguiente, no se trata de una decisión voluntaria de evasión sino de la respuesta del sistema a un mercado formal que no es capaz de absorber a la totalidad de la fuerza laboral disponible bajo las condiciones legales vigentes (Stiglitz, 1986).

2.1.2 Teorías de la informalidad

El análisis académico de la informalidad se basa en el modelo dual de Lewis (1954), quien plantea la existencia de una economía dividida en dos sectores muy diferenciados: uno capitalista dinámico y otro de subsistencia con suministros ilimitados de mano de obra. La tesis central de Lewis es que la informalidad no es un fenómeno permanente, sino una fase transitoria en el proceso de desarrollo económico, en la que el excedente de trabajadores del sector tradicional (normalmente agrícola o de baja tecnología) va siendo absorbido paulatinamente por el sector moderno a medida que éste acumula capital. Sin embargo, este autor advierte que cuando la inversión en el sector formal es insuficiente o el crecimiento poblacional supera la capacidad de creación de empleo técnico, se produce un estancamiento en el sector de subsistencia.

En este escenario, la informalidad continúa siendo una característica estructural en la que el salario se mantiene en niveles mínimos de subsistencia, lo que impide que la economía alcance un equilibrio de pleno empleo formal y perpetúa ciclos de baja productividad. Asimismo, la discusión fue evolucionando con los aportes de Hart (1973), quien luego de sus investigaciones en Ghana introdujo el concepto de “oportunidades de ingresos informales” para dar cuenta de un fenómeno que las estadísticas oficiales no lograban capturar. Hart hizo una importante distinción al señalar que el sector informal no debe entenderse como un mero colectivo de desempleados, sino como un terreno dinámico donde el emprendimiento, el trabajo autónomo y diversas formas de obtener ingresos operan fuera de la regulación estatal. Su postura rompió con la concepción tradicional de la época planteando que la informalidad cuenta con una lógica interna de eficiencia y resiliencia; para Hart las personas ejercen estas actividades no solo por necesidad, sino por la flexibilidad y autonomía que ofrecen frente a la rigidez del mercado laboral asalariado. Esta perspectiva permite entender la informalidad como una forma alternativa de organización

económica, que surge donde el Estado y el mercado formal han fallado en proveer seguridad económica.

Además de esta visión, Harris y Todaro (1970) desarrollan un marco analítico que pretende explicar la persistencia de la informalidad a partir de la teoría de la utilidad esperada y de las distorsiones salariales. Los autores plantean que, si el salario mínimo del sector urbano formal está por encima del equilibrio del mercado, entonces la oferta de trabajo es mayor que la que el sistema puede absorber. Por ello, muchos trabajadores que no consiguen empleo formal se quedan en las zonas urbanas y terminan integrándose en actividades informales. Para Harris y Todaro la informalidad opera como un mecanismo donde los individuos aceptan condiciones laborales precarias y sin seguridad social con la premisa de que estar en el sector informal es el costo de oportunidad necesario para acceder a una vacante en el sector formal. De este modo, la informalidad no es una elección voluntaria de preferencia por el ocio, sino un estado de subempleo estructural con fuerte competencia por ingresos mínimos y riesgo permanente de pobreza. Finalmente, la teoría se consolida con los determinantes institucionales y la medición de la “economía sombra” planteada por Schneider (1986). Este enfoque hace que el debate se centre, ya no en la estructura productiva, sino en el marco legal y tributario del país. Schneider dice que la informalidad es vista como una respuesta racional de los agentes económicos a un sistema que ellos perciben como injusto o ineficiente. Con impuestos elevados, trámites excesivos y servicios públicos deficientes, muchos pequeños negocios y personas encuentran incentivos para permanecer en la informalidad. Schneider logra cuantificar, a través del método de la demanda de efectivo, cómo una mayor presión fiscal hace que una mayor parte del PIB se desvíe hacia la informalidad. Por eso, desde este punto de vista, la informalidad no se explica solamente por la falta de capital o de educación, sino también por las limitaciones institucionales y regulatorias del

estado. Por ello, muchas personas eligen la informalidad como una forma de mantenerse económicamente activas ante normas y requisitos difíciles de cumplir.

2.1.3 Teoría del Capital Humano

La Teoría del Capital Humano explica cómo la capacitación de las personas afecta su rendimiento laboral. Esta teoría propuesta por Schultz en 1961 y sistematizada posteriormente por Becker (1964), señala que los individuos hacen inversiones en sí mismos, principalmente mediante la educación y la capacitación, con la finalidad de aumentar sus conocimientos, habilidades y salud. La educación no es sólo un bien de consumo, sino una inversión que aumenta el valor económico del individuo a largo plazo. Esta idea se basa en la premisa clave de que tener más educación conduce a una mayor productividad marginal. La educación, dice Becker, da al individuo habilidades cognitivas y técnicas que le permiten resolver problemas complejos, adaptarse a nuevas tecnologías y ejecutar procesos organizativos de manera más eficiente que un individuo sin educación formal.

El siguiente paso en esta lógica es que una mayor productividad implica mejores salarios en el mercado laboral: Cuando hay competencia económica, las empresas están dispuestas a pagarle al trabajador de acuerdo con el valor que aporta al proceso productivo. Mincer (1974), señala que, al tener más poder de negociación, una persona con un nivel alto de capital humano puede aspirar a salarios más altos y su contribución al ingreso total de la empresa es mayor. Un título educativo garantiza que el individuo posee las competencias requeridas para desempeñar tareas de gran valor, lo cual incrementa automáticamente el precio de su mano de obra, situándolo en los niveles más altos de la escala salarial.

En consecuencia, la evidencia sugiere que la alta productividad y los mejores salarios hacen más probable la contratación formal. Eso resulta bastante relevante para Colombia, donde la estructura de costos laborales cumple un papel crítico. En Colombia, para que el ser formal signifique que la productividad marginal del trabajador es suficiente para cubrir no solo el salario neto, sino también las cargas prestacionales, aportes a salud, pensiones y otros costos, que constituyen un gasto considerable para las empresas. Si el trabajador tiene bajos niveles de capital humano, su productividad podría ser menor que el costo total de su formalización. Esto genera un incentivo para que el empleador lo mantenga en la informalidad o para que el individuo se vea obligado a buscar medios de subsistencia. Por otro lado, tener un nivel educativo más alto garantiza una productividad que sobrepasa el umbral de los costos legales en Colombia y permite a la persona acceder a contratos protegidos. Así, la educación se convierte en el elemento clave para romper el ciclo de la informalidad, pues mientras mayor sea la cualificación de un trabajador, mayores serán las posibilidades de que el mercado laboral formal quiera incorporarlo.

2.1.4 La Informalidad en el Contexto de la Migración

La migración y la informalidad laboral se basan en la teoría de la segmentación de los mercados de trabajo y la búsqueda de medios de subsistencia inmediatos ante la ausencia de redes de apoyo formales en el país de destino. De acuerdo con la perspectiva de Piore (1979) sobre la Teoría del Mercado Dual de Trabajo, las economías de acogida tienden a desarrollar un sector secundario con empleos de baja cualificación, inestables y sin seguridad social, que es en donde se emplea mayoritariamente a la población migrante. En este sentido, la informalidad no es sólo un producto de la inexistencia de documentos legales, sino una característica intrínseca de ciertos sectores económicos que necesitan mano de obra flexible y barata para seguir siendo competitivos.

Para el migrante, el sector informal es la vía más fácil de acceso al mundo laboral, ya que le permite obtener ingresos desde su primer día de llegada, aunque esto signifique aceptar condiciones precarias y renunciar a derechos prestacionales básicos.

Portes y Haller (2004) plantean, en el marco de un análisis que acompaña este análisis, que la informalidad tiende a sostenerse a través de redes sociales y conexiones de capital social étnico en contextos migratorios. Estas redes permiten a los migrantes insertarse en nichos económicos específicos –como el comercio ambulante, los servicios domésticos o la construcción– que se rigen por lógicas de reciprocidad y no por contratos legales. Este refugio económico, sin embargo, tiene un carácter dual, ya que al mismo tiempo que actúa como un mecanismo de protección social frente a la exclusión del Estado, también genera una trampa de pobreza donde el migrante queda limitado a actividades de baja productividad y con pocas posibilidades de mejorar sus condiciones económicas. En ciudades muy pobladas donde las instituciones son débiles, como es el caso de las fronteras, cuando llega un gran número de trabajadores migrantes, también pueden disminuir los ingresos en el sector informal, ya que crece la competencia entre los trabajadores nacionales y los extranjeros por utilizar los espacios públicos y los recursos disponibles en las actividades económicas que se realizan en la calle.

Finalmente, la teoría de la “asimilación segmentada” de Zhou (1997), advierte que una prolongada permanencia en la informalidad puede dar lugar a procesos de integración descendente por los que los migrantes acaban insertándose en los estratos más bajos de la estructura social del país receptor. El hecho de no regularizarse como migrante constituye una barrera que empuja al individuo hacia donde el riesgo de explotación laboral es significativamente mayor. Desde este punto de vista, la irregularidad migratoria no debe verse como un problema de voluntad individual, sino como una consecuencia de la desconexión entre las políticas de control fronterizo y las

realidades de un mercado laboral que exige trabajadores, pero que no está dispuesto ni capacitado para regularizarlos. De ese modo, la migración emerge como un factor que ensancha la base de la pirámide informal, cambiando las dinámicas urbanas, planteando retos críticos para la política pública y la cohesión social.

2.1.5 Contexto estructural de la informalidad laboral en Colombia

Desde hace mucho, en el mercado laboral colombiano se puede observar una considerable segmentación. El sector informal emplea a quienes no encuentran cabida en el sector formal. Es por razón de rigideces institucionales y bajos niveles de industrialización. La crisis del COVID-19 provocó un cambio radical en esta estructura. La ocupación se contrajo sin precedentes a finales de 2021, efecto que fue registrado por el Departamento Administrativo Nacional de Estadística (2021). A nivel nacional, la informalidad alcanzó un nivel del 58,1%, lo que significa que durante dicho período de reactivación inmediata más de la mitad de la fuerza de trabajo no contaba ni con protección social, ni con estabilidad contractual.

A nivel regional la informalidad muestra brechas estructurales, que dividen al país en función de su capacidad productiva y ubicación. Según el DANE (2021), aunque en las ciudades con mayor densidad empresarial como Bogotá (41,4%) y Medellín (40,8%) la informalidad se presentaba por debajo del promedio nacional, los centros urbanos mostraban cifras alarmantes. En 2021, ciudades como Sincelejo (67,6%) y Cúcuta (71,5%) registraron las tasas más altas de ocupación informal, lo que refleja una dependencia casi total de actividades de baja productividad y del comercio minorista. Esta desigualdad se mantuvo hasta el primer trimestre de 2024, cuando ciudades como Manizales (33,4%) mostraron tasas que son menos de la mitad de las de Sincelejo

(68,5%) o Cúcuta (61,1%), mostrando que la informalidad en Colombia es un fenómeno con fuerte carga territorial, según el DANE (2024).

En 2024 continuó, pese a cierta estabilidad macroeconómica. De acuerdo con el DANE (2024), la tasa de informalidad a nivel nacional para el trimestre diciembre de 2023 a febrero de 2024, se ubicó en 55,8 %. En este sentido, la ocupación por cuenta propia continúa siendo la modalidad más frecuente, involucrando a más del 40 % de los ocupados. De igual manera, la permanencia de estas cifras indica que la informalidad no es un estado coyuntural, sino un rasgo sistémico. La brecha de género lo muestra de manera muy clara, en las 23 principales ciudades la tasa de informalidad de las mujeres era de 45,3% a principios de 2024 superando el 42,6% de los hombres. Esa realidad nos muestra que, en departamentos fronterizos como el Norte de Santander, la informalidad es la respuesta a la escasez de un tejido industrial capaz de competir con economías de subsistencia, lo cual exige un análisis diferenciado de los factores sociodemográficos entre 2021 y 2024.

2.1.6 Contexto estructural de la informalidad laboral en Norte de Santander

En el departamento de Norte de Santander, en su capital Cúcuta, existe una curiosa situación de informalidad laboral en Colombia, lo cual se debe a que se trata de una zona fronteriza. Desde siempre la economía de la región ha dependido de los intercambios comerciales con Venezuela. Eso ha dado lugar a un mercado laboral que se sustenta en el intercambio de bienes y servicios de baja escala, muchas veces fuera de la regulación estatal. Según datos del Departamento Administrativo Nacional de Estadística (2021), durante la etapa de reactivación postpandemia del año 2021, Cúcuta fue de forma constante la ciudad con mayor tasa de informalidad del país, superando el 70%. Esta cifra es una señal de debilidad productiva local, donde la falta de una sólida

industria manufacturera y la dependencia del diferencial cambiario han llevado a que la mayoría de la gente tenga que buscar el “rebusque”, el comercio informal, como única forma de subsistir.

La pandemia en Cúcuta golpeó más fuerte sobre las vulnerabilidades que ya estaban allí. «Al contrario de otras capitales colombianas con economías más diversificadas, el cierre de la frontera y las medidas de confinamiento paralizaron también la informalidad de la ciudad, dejando a miles de familias sin ingresos por la ausencia de redes de protección social». Entre 2021 y 2022, la informalidad no solo se mantuvo en niveles elevados, sino que se convirtió en una estrategia de subsistencia ante la llegada de nuevos trabajadores al mercado laboral. En Cúcuta, la informalidad se debe a una “cultura de frontera”, en la que el cruce de la frontera e la inestabilidad política de Venezuela generan un clima de incertidumbre que desincentiva la inversión formal a largo plazo, lo que perpetúa ciclos de ocupación precaria y baja productividad. Todo esto también se debe al fuerte impacto que ha tenido la migración en los últimos años.

Para el 2024, las dinámicas laborales en Norte de Santander demuestran que, aún con la reapertura gradual de los puentes internacionales, la informalidad no se ha reducido de manera significativa. De acuerdo con el DANE (2024), en la ciudad la mayor parte de la mano de obra informal se concentra en sectores como el comercio minorista, el transporte y los servicios hoteleros. La persistencia de tasas de informalidad que superan el 65% en el 2024, pone de relieve que la problemática de Cúcuta no es solo temporal, sino que está muy relacionada con la falta de cualificación del capital humano y la ausencia de políticas públicas integrales que conecten a la región con los mercados del interior del país. De este modo, el mercado laboral nortesantandereano se muestra como un escenario de intensa segmentación, en el que el empleo formal es privilegio de pocos, mientras la mayor parte de la población ocupada vive en la desprotección social, lo que representa un mayor desafío para la cohesión socioeconómica del departamento.

2.2 Marco conceptual

La informalidad laboral se ha analizado desde distintos enfoques, debido a la complejidad de ésta. La economía informal está formada por actividades económicas llevadas a cabo por trabajadores y unidades productivas que carecen de una cobertura suficiente de la legislación laboral y de los sistemas de protección social (OIT, 2002). En este sentido, la informalidad no está asociada solamente a la falta de afiliación a la seguridad social, sino también a condiciones laborales precarias y a un acceso limitado a los derechos laborales. Asimismo, la OIT (2015) considera que la formalización laboral es un proceso de carácter multidimensional, que implica la transición de los trabajadores y las empresas al cumplimiento de las normas legales, tributarias y de seguridad social, y es uno de los principales objetivos de política pública de los países en desarrollo.

La oferta y la demanda de mano de obra dentro del mercado laboral determinan los niveles de ocupación y los salarios de una economía (Samuelson y Nordhaus, 2010). Dentro de esta dinámica, estar ocupado significa que una persona en edad laboral generó algún tipo de valor o producción para recibir un pago, tal como lo indica la OIT (2013). En Colombia, para medir estas variables se tiene la GEIH, a través de la encuesta, que nos permite conocer la realidad de la formalidad e informalidad de Colombia.

El capital humano es uno de los principales determinantes individuales de la informalidad. Becker (1993), sostiene que la educación, la formación y la experiencia son inversiones que aumentan la productividad y los ingresos potenciales de los individuos. Por consiguiente, los niveles de educación más elevados reducen la probabilidad de ser insertados en empleos informales, ya que amplían las posibilidades de acceder a ocupaciones mejor remuneradas y formales. La Clasificación Internacional Normalizada de la Educación (UNESCO, 2012), permite

agrupar el nivel educativo alcanzado por los individuos, variable empleada como sustituto del capital humano en estudios económicos.

Variables demográficas como la edad influyen en la dinámica del mercado de trabajo. La edad es el ciclo de vida productivo y la acumulación de experiencia. Factores que influyen en la estabilidad laboral. Varios estudios muestran que la informalidad tiende a concentrarse en las personas en edades tempranas y avanzadas, afectando especialmente a jóvenes con poca experiencia y a adultos mayores que encuentran barreras para su inserción formal. De la misma manera, la condición migratoria es un factor determinante, ya que los migrantes pueden tener restricciones legales, dificultades en el reconocimiento de títulos y limitaciones en sus redes de apoyo, lo que aumenta su probabilidad de participar en actividades informales (Organización Internacionales de las Migraciones [OIM], 2019).

Desde el punto de vista metodológico el análisis de la informalidad a través de una variable independiente dicotómica (Formal/Informal) requiere el uso de modelo de elección binaria. La regresión logística es una técnica empleada para calcular la probabilidad de que un determinado evento ocurra en función de varias variables explicativas como lo plantean Hosmer et al. (2013). El modelo logit utiliza una función logística que asegura que las probabilidades calculadas se sitúen dentro del rango de 0 a 1, tal como señalan Gujarati y Porter (2010). Debido a que el modelo es no lineal, los coeficientes estimados no se interpretan directamente como cambios en la probabilidad, por lo tanto, se usan efectos marginales para medir la variación en la probabilidad de informalidad ante cambios en las variables independientes (Greene, 2018).

Estos conceptos, tomados en conjunto, facilitan la estructuración del análisis de los determinantes sociodemográficos y laborales de la informalidad en Norte de Santander,

articulando el marco teórico con la estrategia econométrica empleada, y le dan coherencia al estudio.

2.3 Marco legal

El trabajo es un derecho fundamental reconocido por el sistema jurídico colombiano y uno de los pilares esenciales del estado social de derecho. En la Constitución Política de Colombia (1991), en su artículo 25, se estipula que el trabajo está bajo la especial protección del Estado y toda persona tiene derecho a condiciones dignas y justas. El artículo 48 señala, por otra parte, que la seguridad social es un servicio público obligatorio y el artículo 53 establece principios como la igualdad de oportunidades y la remuneración mínima vital para los ciudadanos. Estos mandatos conforman la base sobre la cual se construye la regulación laboral en el país y establecen los cimientos de la formalidad.

La ley 100 de 1993 creó el sistema de seguridad social integral y constituye un referente central para el estudio de la informalidad, ya que la afiliación a salud, pensiones y riesgos laborales es el criterio técnico primordial para diferenciar el empleo formal del informal. Por su parte la ley 789 de 2002 y la ley 429 de 2010 (Ley de formalización y generación de empleo) han buscado dinamizar el mercado mediante incentivos tributarios y parafiscales especialmente relevantes para microempresas que predominan el departamento de Norte de Santander.

En los territorios fronterizos, como el estudiado, se hace de gran importancia el Decreto 216 de 2021, creado a través del Estatuto temporal de protección para migrantes venezolanos (ETPV), instrumento que permite el Permiso de Protección Temporal (PPT), el cual se convirtió en un documento de identificación válido para la contratación legal y la afiliación al sistema de

seguridad social. Un cambio técnico fundamental fue facilitar la bancarización del migrante, para que pudiese abrir cuentas para recibir su nómina.

3. Método

El presente trabajo se desarrolla desde un enfoque cuantitativo, apoyado en técnicas estadísticas y econométricas orientadas al análisis de los determinantes de la informalidad laboral en Norte de Santander (2021 y 2024). La metodología empleada utiliza los microdatos de la Gran Encuesta Integrada de Hogares (GEIH) del DANE, lo cual permite caracterizar las condiciones laborales y los factores asociados a la probabilidad de estar en empleo informal.

La investigación adopta un diseño no experimental transversal para cada año, lo que permite realizar comparaciones temporales. Individuos ocupados de 15 años y más (DPTO = 54) son la unidad de análisis. El estudio se limita exclusivamente a la población ocupada, para evitar sesgos de selección, bajo el supuesto de que la informalidad (entendida como el ejercicio de actividades económicas sin protección de seguridad social) es una condición propia de los que participan efectivamente en el mercado laboral, independientemente de la percepción de un salario.

Metodológicamente, la información se procesó integrando los módulos de características generales, ocupados, migración y vivienda de la GEIH. Para ambos años se construyó de forma consistente la variable de migración a partir del lugar de nacimiento (“nacido en otro país”), identificando esta condición como una variable exógena clave.

La variable dependiente, informalidad laboral, se aproximó mediante un proxy basado en la no cotización al sistema de pensiones. Esta definición, aunque no coincide con la metodología oficial del DANE (que incluye criterios adicionales del tamaño de la empresa y la posición

ocupacional), se emplea de forma generalizada en la literatura por su disponibilidad y consistencia en la GEIH.

Entre las variables explicativas se encuentran: edad, sexo, nivel de educación, ocupación (categoría ocupacional), estrato socioeconómico y migración. Los modelos se estimaron incorporando el complejo diseño muestral de la GEIH (factor de expansión FEX_C18, UPM y estrato) para garantizar la validez estadística. Para la interpretación, se calcularon Efectos Marginales Promedio (AME) y probabilidades predichas. Por último, se realizó una prueba de robustez estimando un modelo alternativo en el que la informalidad se define a partir de la no afiliación al sistema de salud, evaluando la estabilidad de los resultados frente a distintas aproximaciones. Los tests de Wald para la significancia conjunta de las variables confirmaron la relevancia de los modelos.

En las diferentes etapas metodológicas se utilizó el apoyo puntual de herramientas de inteligencia artificial (Elicit para revisión de literatura y Microsoft 365 Copilot para apoyo en redacción), empleándose estrictamente como soporte complementario a la investigación.

3.1 Modelo Econométrico

Para el análisis econométrico se utilizó un modelo de regresión logística (Logit), técnica apropiada debido a que la variable dependiente es dicotómica, codificándose como 1 si el individuo es informal y 0 si es formal. Para aproximar la condición de informalidad laboral, se utiliza un proxy basado en la no contribución al sistema de pensiones. Aunque la medición oficial del DANE incorpora un conjunto más amplio de variables, este acercamiento está ampliamente avalado por la literatura para capturar de forma consistente la inserción laboral desde la perspectiva de la protección social efectiva.

La definición técnica tiene un sustento académico: Según Bernal y otros autores, las aportaciones a la seguridad social constituyen una medida razonable de informalidad, puesto que permiten determinar si la persona dispone del paquete completo de prestaciones propias de la ocupación formal. En esa misma línea, Galvis define técnicamente la falta de afiliación a salud y pensión. Aunque esta métrica no captura otras dimensiones, como el tamaño de la empresa, permite un análisis robusto de la desprotección social, criterio que en la literatura nacional ha reportado niveles de informalidad cercanos al 60% en contextos urbanos.

El modelo incluye variables explicativas sociodemográficas y laborales como: edad, sexo, nivel educativo, categoría ocupacional, estrato socioeconómico y migración. En particular, la variable de migración se construyó únicamente a partir del lugar de nacimiento (“nacido en otro país”). La decisión metodológica está en función de las limitaciones propias de la GEIH, que no proporciona información sistemática ni continua sobre estatus legal o permisos de trabajo para los años analizados. Por lo tanto, el lugar de nacimiento es la única variable exógena estable y comparable, que sirve como proxy para analizar la inserción de la población migrante en el mercado laboral fronterizo y evaluar, de manera indirecta, el impacto de los procesos institucionales de integración entre 2021 y 2024. Debemos aclarar que la variable categoría ocupacional se refiere a la posición en la ocupación del individuo, y que debe diferenciarse de la limitación metodológica de trabajar sólo con la población ocupada, utilizada como filtro de selección de la muestra.

Finalmente, el estrato socioeconómico se incluyó como una variable categórica construida a partir de la variable P4030S1A1 de la GEIH. Esta se recodificó en niveles, tomando como categoría de referencia el estrato más bajo, con el propósito de analizar las diferencias relativas en la probabilidad de informalidad entre los distintos estratos socioeconómicos.

La especificación funcional del modelo se expresa de la siguiente forma linealizada (log-odds)

$$\ln\left(\frac{P(\text{Informal}_i = 1)}{1 - P(\text{Informal}_i = 1)}\right) \\ = \beta_0 + \beta_1 \text{Migrante}_i + \beta_2 \text{Edad}_i + \beta_3 \text{Sexo}_i + \beta_4 \text{Educación}_i + \beta_5 \text{Ocupado}_i \\ + \beta_6 \text{Estrato socioeconómico}_i + \varepsilon_i$$

Donde:

$P(\text{Informal}_i = 1)$ Probabilidad de que el individuo i sea informal.

β_0 Intercepto del modelo.

$\beta_1, \dots, \beta_6$ Coeficientes asociados a las variables explicativas.

ε_i Término de error.

Al emplear el modelo Logit, se asegura que las predicciones sigan dentro del rango lógico de probabilidad [0,1], evitando las limitaciones del modelo lineal de probabilidad. Por otra parte, esta visión se basa en la teoría de la utilidad aleatoria, que asume que los individuos toman decisiones de trabajo bajo restricciones estructurales, donde la informalidad puede ser entendida como un resultado de equilibrio entre las condiciones individuales y las oportunidades del mercado.

Como los coeficientes de un modelo Logit no pueden interpretarse directamente como cambios en la probabilidad, se calcularon los efectos marginales promedio (AME). Esto hizo más fácil y clara la interpretación económica de los resultados, pues permitió presentarlos en función de la variación de la probabilidad de informalidad cuando cambian las variables explicativas. Adicionalmente, se construyeron gráficos del comportamiento de estos determinantes basados en las probabilidades predichas por el modelo, facilitando la identificación visual de brechas y diferencias en la probabilidad de informalidad entre los diferentes grupos de la población.

Para estimar el modelo se empleó el complejo diseño muestral de la GEIH. Se introdujo en este proceso el factor de expansión (FEX_C18), la unidad primaria de muestreo (UPM) construida a partir de las variables Directorio y Secuencia. De esta forma se pretendió asegurar que los resultados fueran representativos de la población ocupada del departamento de Norte de Santander (DPTO = 54). También se probó la significancia global del modelo mediante la prueba de Wald y se hicieron análisis de robustez con una medida alternativa de informalidad definida por la afiliación al sistema de salud tanto para el año 2021 como para el año 2024.

Finalmente, se realizó el análisis estimando modelos separados para los años 2021 y 2024, lo que permite identificar cambios en la magnitud y dirección de los determinantes de la informalidad en el tiempo, manteniendo la coherencia metodológica entre ambos periodos.

3.2 Metodología y Proceso de Análisis en Modelo

Para realizar el análisis, se emplearon herramientas estadísticas y modelos econométricos aplicados a los datos de la GEIH en los Software Stata y R.

3.2.1 Variables Clave

Debido a la actualización metodológica de la Gran Encuesta Integrada de Hogares (GEIH), la variable migración presentan diferencias en su nomenclatura entre los años 2021 y 2024. No obstante, es importante resaltar que las variables incluidas en el análisis son conceptualmente equivalentes en ambos periodos, lo que garantiza la comparabilidad de los resultados. La única diferencia relevante se presenta en la variable asociada al lugar de nacimiento (migración), cuya denominación cambia entre años, aunque mide exactamente el mismo concepto.

Tabla 1. *Presentación de variables para cada año*

Variables 2021	Variables 2024
P6920: ¿Está cotizando actualmente a un fondo de pensiones?	P6920: ¿Está cotizando actualmente a un fondo de pensiones?
P6040: ¿Cuántos años cumplidos tiene?	P6040: ¿Cuántos años cumplidos tiene?
P3271: Sexo	P3271: Sexo al nacer
P3042: Nivel educativo alcanzado	P3042: Nivel educativo alcanzado
OCI: Población ocupada	OCI: Población Ocupada
P6430: En este trabajo ... Es:	P6430: En este trabajo ... Es:
P756: ¿Dónde nació?	P3373: ¿Dónde nació?
P4030S1A1: Estrato por tarifa	P4030S1A1: Estrato por tarifa
Prueba De Robustez	
P6090: ¿... Está afiliado, es cotizante o es beneficiario de alguna entidad de seguridad social en salud?	P6090: ¿... Está afiliado, es cotizante o es beneficiario de alguna entidad de seguridad social en salud?
P6100: ¿A cuál de los siguientes regímenes de seguridad social en salud está afiliado:	P6100: ¿A cuál de los siguientes regímenes de seguridad social en salud está afiliado:
P6110: ¿quién paga mensualmente por la afiliación de ...?	P6110: ¿quién paga mensualmente por la afiliación de ...?

- Variable Dependiente

Informalidad laboral: Variable dicotómica que identifica la condición del trabajador (1 = informal; 0 = formal). Se construyó a partir de la no cotización a pensión (P6920), utilizada como proxy de informalidad. Si bien esta medida no corresponde exactamente a la metodología oficial del DANE, constituye una aproximación ampliamente utilizada en la literatura, permitiendo capturar la inserción laboral informal desde la perspectiva de la protección social. Con la intención de profundizar en la naturaleza de la informalidad, se añadieron a las pruebas de robustez las variables P6100 y P6110. Esta combinación permite un análisis más fino, ya que permite diferenciar entre la cobertura general y la calidad del aporte (diferenciando entre trabajadores dependientes e independientes). La variable P6440 también fue incluida en ejercicios de contraste

para comprobar la estabilidad de los resultados frente a distintas dimensiones de la desprotección laboral.

- Variables Independientes

Edad: Variable continua medida en años cumplidos (P6040), utilizada para aproximar la experiencia laboral y el ciclo de vida del individuo.

Sexo: Variable categórica que distingue entre hombres y mujeres. Para 2021 se utiliza la variable P6020 y para 2024 la variable P3271 (sexo al nacer), manteniendo equivalencia conceptual entre ambas.

Migración: variable dicotómica construida a partir del lugar de nacimiento del individuo. Para garantizar la comparabilidad entre años, se realizó una homologación de la variable, utilizando P756 en 2021 y P3373 en 2024, ambas referidas a la pregunta “¿dónde nació?”. A partir de esta información se definió la variable migrante, donde toma el valor de 1 si el individuo nació en otro país y 0 en caso contrario. Esta construcción asegura consistencia metodológica entre las dos bases de datos. Se escogió esta variable con el fin de garantizar la consistencia intertemporal y la comparabilidad de los microdatos. Aunque el estatus legal y los permisos de trabajo son factores determinantes críticos en la literatura más reciente, la falta de un identificador sistemático de legalidad en la GEIH no permite su inclusión directa. El lugar de nacimiento se utiliza entonces como variable proxy de la condición migratoria, de la cual se discuten en profundidad sus alcances y limitaciones en la sección de discusión del documento."

Educación: variable categórica reclasificada en tres niveles (baja, media y alta). Esta clasificación permite capturar diferencias en capital humano y productividad laboral.

Ocupación: esta variable se construyó con base en la información de la posición ocupacional del individuo (P6430) de la GEIH, respetando el filtro anterior de ocupación (OCI=1).

Se procedió a una reclasificación técnica de las categorías originales para ajustar los datos a la estructura del modelo econométrico, consolidándolas en tres niveles analíticos: Asalariado (empleados privados y gubernamentales), Independiente y Precario/Otros. Este agrupamiento estratégico permite captar con mayor precisión las diferencias estructurales en la inserción laboral y su relación con la informalidad, disminuyendo la dispersión estadística y garantizando la veracidad de los coeficientes estimados en el modelo.

Estrato socioeconómico: categórica variable construida a partir de la variable original P4030S1A1 de la GEIH que clasifica los hogares de acuerdo con el nivel socioeconómico. Con el fin de mejorar la robustez estadística y asegurar una adecuada representación de las observaciones en la muestra regional, la variable se reclasificó en dos niveles: Estrato Bajo” (agrupando los estratos 1 y 2) y “Estrato Medio-Alto” (agrupando los estratos 3 al 6). Esta variable fue incorporada al modelo como factor (estrato_soc), tomando como categoría de referencia el “Estrato Bajo”. Esta estructura permite hacer un análisis comparativo de las brechas en la probabilidad de informalidad entre los hogares que reciben subsidios a servicios públicos y aquellos que tienen una mayor capacidad contributiva, lo que facilita una interpretación más clara del impacto del estrato socioeconómico en la inserción laboral formal.

3.2.2 Preparación de la Base de Datos

Los datos se procesaron por separado para los años 2021 y 2024, integrando en un principio los módulos de la Gran Encuesta Integrada de Hogares (GEIH) utilizando características generales, ocupados, vivienda y migración, mediante procesos de fusión (merge) y unión vertical (append) para armar una base anual representativa de Norte de Santander correspondiente a cada año. Posteriormente, los microdatos se transfirieron al entorno estadístico R donde se realizó un

riguroso protocolo de limpieza, estandarización y depuración. En esta etapa se aplicaron filtros estructurales determinantes: se restringió la muestra a individuos de 15 años y más y se utilizó la variable OCI (Ocupados) con valor 1, para asegurar que el análisis se realizara exclusivamente sobre la población efectivamente ocupada, evitando sesgos derivados de la inclusión de individuos inactivos o desocupados. De igual manera, se hizo la homologación de variables clave como la condición migratoria (P756 para 2021 y P3373 para 2024), nivel educativo y estrato socioeconómico. En especial, la variable de posición ocupacional (P6430) fue sometida a un proceso de reclasificación técnica, a través de una estructura de agrupamiento que condensó las categorías originales en tres niveles analíticos: Asalariado (incluyendo empleados privados y del gobierno), Independiente y Precario/Otros, construyendo así una taxonomía robusta que mitiga la dispersión estadística y elimina categorías poco representativas. Para asegurar la calidad de los estimadores se procedió a una estricta limpieza que consistió en la eliminación sistemática de todas las observaciones con valores perdidos (NA) tanto en las variables del modelo, como en los factores de expansión (FEX_C18) y los identificadores de diseño muestral. Por último, se estableció el diseño muestral complejo mediante la interacción de las variables Directorio y Secuencia para definir la Unidad Primaria de Muestreo (UPM), aplicando el ajuste para las UPM solitarias y validando la representatividad poblacional. Al finalizar este largo proceso, la muestra quedó conformada por 10.974 observaciones para el año 2021 y 10.808 para el año 2024, que constituyen el insumo robusto para estimar los modelos logísticos mediante la función `svyglm`.

3.2.3 Construcción de variables del modelo principal

Con el propósito de optimizar la estimación del modelo econométrico y facilitar la interpretación de los resultados, se realizaron transformaciones y reclasificaciones sobre las

variables originales de la GEIH, garantizando consistencia metodológica entre los años 2021 y 2024:

Variable de informalidad (informal): Se definió como una variable dicotómica (0–1), donde 1 indica que el individuo se encuentra en condición de informalidad y 0 en condición de formalidad. Esta variable se construyó a partir de la no cotización a pensión (P6920), utilizada como proxy de informalidad, en línea con la disponibilidad de información en la GEIH.

Variable de sexo (sexo): Se incorporó como una variable categórica, donde el valor 1 corresponde a “Hombre” (categoría de referencia) y el valor 2 a “Mujer”.

Variable de educación (educ_level): Se construyó a partir de los niveles educativos originales, reclasificándolos en tres categorías: “Baja”, “Media” y “Alta”. Esta agrupación permite capturar diferencias en capital humano y facilitar la interpretación de los efectos dentro del modelo.

Variable de migración (migrante): Se construyó como una variable dicotómica a partir del lugar de nacimiento del individuo, utilizando P756 para 2021 y P3373 para 2024. Se definió como migrante a quienes reportan haber nacido en otro país (valor 1) y como no migrante a los nacidos en Colombia (valor 0), garantizando consistencia entre ambas bases.

Variable de ocupación (ocupacion): Esta variable se construyó con base en la clasificación de posición ocupacional reportada en la GEIH (variable P6430). La encuesta enumera inicialmente diversas formas de vincularse al trabajo (obrero/empleado particular o del gobierno, trabajador independiente, patrón o empleador, trabajador familiar sin remuneración, trabajador en otra modalidad, entre otros). Para este análisis se recodificó en tres categorías analíticas: “Asalariado” (incluye a los obreros o empleados del sector público y privado), “Independiente” (agrupa a los trabajadores por cuenta propia, patrones o empleadores) y “Precario/Otros” (integra a los trabajadores familiares sin remuneración y otras formas de ocupación sin vinculación formal). Esta

clasificación permite segmentar el mercado laboral según la naturaleza del vínculo, tomando como referencia al “Asalariado”, por ser el grupo que bajo la tipología de la encuesta mantiene una relación de subordinación laboral, lo que facilita comparar de qué manera estas modalidades inciden en la probabilidad de acceso a la protección social.

Variable de estrato socioeconómico (estrato_soc): Variable construida a partir de la variable original P4030S1A1 de la Encuesta de ingresos y gastos de los hogares. Esta variable se recodificó a dos niveles: Estrato Bajo (en el que se agruparan los estratos 1 y 2) y Estrato Medio Alto (en el que se agruparan los estratos 3 al 6). Este agrupamiento se hizo con la finalidad de dar robustez estadística y asegurar una adecuada representación en la muestra, ya que la baja frecuencia de observaciones en los estratos más altos en la región, podría comprometer la estabilidad y significancia de los estimadores en el modelo logístico. Al incorporar esta variable como factor (estrato_soc) y establecer al “Estrato Bajo” como referencia, el modelo estima el efecto diferencial sobre la probabilidad de informalidad, lo cual permite capturar con mayor precisión las disparidades socioeconómicas en la inserción laboral regional.

3.2.4 Especificación del Modelo de Probabilidad Logit

Para identificar los determinantes de la informalidad laboral en el departamento de Norte de Santander, se estimó un modelo de regresión logística (Logit), apropiado para el análisis de variables dependientes dicotómicas. Este modelo permite estimar la probabilidad de que un individuo ocupado se encuentre en condición de informalidad en función de un conjunto de variables explicativas, garantizando que las predicciones se mantengan dentro del rango [0,1].

A diferencia del modelo lineal de probabilidad, el modelo Logit captura la naturaleza no lineal de la relación entre las características socioeconómicas del individuo y su condición laboral. En este estudio, la especificación del modelo incorpora simultáneamente variables demográficas

(edad y sexo), de capital humano (nivel educativo), de inserción laboral (ocupación), de contexto socioeconómico (estrato) y de migración (lugar de nacimiento).

Adicionalmente, la estimación se realizó incorporando el diseño muestral complejo de la GEIH, incluyendo el factor de expansión (FEX_C18), las unidades primarias de muestreo (UPM) lo que permite obtener estimaciones representativas y errores estándar robustos. El modelo fue estimado de manera independiente para los años 2021 y 2024, con el fin de comparar los determinantes de la informalidad en ambos periodos.

3.2.5 Estimación de Efectos Marginales Promedio (AME) – Probabilidades Predichas.

Dado que los coeficientes estimados en el modelo Logit se interpretan en términos de log-odds (logaritmo de la razón de probabilidades), su interpretación directa no resulta intuitiva desde el punto de vista económico. Por esta razón, el análisis se complementa con el cálculo de los Efectos Marginales Promedio (Average Marginal Effects, AME), los cuales permiten expresar los resultados en términos de cambios en probabilidad.

Se estimaron los efectos marginales promedio (AME) a partir del modelo logit con herramientas estadísticas en R, lo que permitió medir el efecto promedio que cada variable explicativa tiene sobre la probabilidad de pertenecer al empleo informal, manteniendo constantes las demás características incluidas en el modelo.

Esto permite expresar los resultados en porcentajes, lo que facilita su interpretación y comparación entre grupos y periodos de análisis. Es importante destacar que los efectos marginales permiten conocer la relación entre algunas características y la probabilidad de estar en la informalidad laboral. Sin embargo, los resultados no deben ser interpretados como evidencia de

que una variable causa directamente la otra. Reflejan más bien tendencias o comportamientos promedio de la población estudiada.

Adicionalmente al cálculo de los efectos marginales, se hicieron gráficos de probabilidades predichas. El objetivo de estas visualizaciones es facilitar la interpretación de las interacciones no lineales del modelo y permitir una comparación intuitiva entre los años 2021 y 2024. A diferencia de los AME, que dan un valor promedio puntual, las probabilidades predichas permiten ver cómo cambia la probabilidad de informalidad ante cambios en variables continuas (como la edad) o categóricas (como el nivel educativo), manteniendo las demás variables en sus medias o valores de referencia.

La razón para realizar estas gráficas es detectar brechas estructurales. Al graficar las probabilidades predichas contra los años de análisis, es posible observar visualmente si las diferencias en la probabilidad de informalidad entre grupos por ejemplo, entre migrantes y no migrantes se han ampliado, contraído o mantenido estables en el periodo de estudio. Este tratamiento gráfico permite leer los resultados de manera dinámica, convirtiendo los estimadores estadísticos en patrones de comportamiento que facilitan la identificación de los focos de vulnerabilidad en el mercado laboral de Norte de Santander, entregando una evidencia visual clara sobre la evolución de la informalidad que va más allá de la interpretación de los coeficientes individuales.

3.2.6 Limitaciones y consideraciones metodológicas.

El presente estudio se desarrolló bajo un riguroso marco metodológico; sin embargo, se reconocen limitaciones inherentes al diseño de investigación que deben ser tenidas en cuenta para una correcta interpretación de los hallazgos:

Alcance a la población empleada: el análisis está circunscrito exclusivamente a la población ocupada, esto es, a aquellas personas que participan efectivamente en el mercado de trabajo. Aunque este límite es necesario para estimar la informalidad, se reconoce que los resultados son específicos para esta población y el análisis se limita a quienes ya tienen una relación laboral. Desde el punto de vista metodológico, la decisión es correcta para el objetivo planteado, puesto que permite identificar los factores que determinan la precariedad de los que ya están en el mercado laboral.

Construcción de la variable dependiente y comparación con el DANE: para medir la informalidad, el DANE utiliza una metodología oficial, multidimensional, que clasifica a los trabajadores de acuerdo con el tamaño de la unidad económica y el cumplimiento de requisitos legales y contables (registro mercantil, contabilidad formal y número de empleados). Por el contrario, en este estudio se tomó la cotización a pensión como proxy de informalidad laboral, con el fin de medir la desprotección social y la precariedad a largo plazo. Para garantizar la robustez de los resultados, se utilizó el acceso a la seguridad social en salud solamente para un ejercicio de robustez manteniendo ambas dimensiones como modelos independientes, para evitar mezclas metodológicas que pudieran comprometer la validez estadística.

Condición de migración: la variable de estatus migratorio se construyó solo a partir del lugar de nacimiento, debido a que los microdatos de la GEIH no incluyen información sobre la regularización administrativa de los migrantes. Se reconoce que esta variable es un indicador de procedencia poblacional y no un indicador directo de la situación legal. También se omitieron variables como la tenencia de contrato escrito para medir el estatus, ya que esto incurriría en un problema de endogeneidad al estar directamente relacionado con la propia definición de formalidad laboral.

4. Resultados

4.1 Tasa de informalidad laboral

Para calcular la tasa de informalidad laboral para Norte de Santander se usaron los factores de expansión y la estructura de diseño muestral complejo de la GEIH.

En 2021, la tasa de informalidad laboral fue del 83,17% (IC 95% [82,26%-84,08%]). En 2024 se determinó que la prevalencia fue del 78,88% (IC95% [77,86%-79,90%]).

Como los intervalos de confianza de los dos años no se superponen, podemos concluir de forma estadísticamente rigurosa que la caída observada de 4,29 puntos porcentuales es consistente y significativa. El hecho de que no se superpongan nos permite asegurar que la caída en los porcentajes de informalidad no es un resultado de la variabilidad muestral, sino que refleja un cambio real en la estructura del mercado laboral departamental. Es importante precisar que estos niveles, superiores a los valores reportados oficialmente por el DANE para áreas metropolitanas, se corresponden con nuestra definición operativa basada en la protección a la seguridad social y en la cobertura departamental total del estudio, la cual incluye la alta precariedad de las zonas rurales y no metropolitanas.

Con esta mejora adicional no se logra bajar los niveles de informalidad, que siguen siendo elevados y relevantes en el entorno regional. Este fenómeno estaría señalando la existencia de barreras estructurales, tales como la baja productividad, el predominio del trabajo independiente y la escasez de empleos formales, que restringen la inserción laboral en el Norte de Santander. En este contexto, la continuidad de la informalidad pone de realce la necesidad de examinar los factores individuales que inciden sobre la probabilidad de incorporación informal, tarea que se aborda en el siguiente apartado a través de la estimación de los modelos econométricos logit.

4.2 Resultados del modelo Logit

Los resultados de los modelos logit para 2021 y 2024, confirman que la informalidad en Norte de Santander, responde a factores estructurales persistentes. El hecho de que todas las variables incluidas sean estadísticamente significativas (en especial la condición migratoria, la educación y el tipo de ocupación) evidencia que el modelo es sólido para explicar la probabilidad de inserción en el sector informal. Los determinantes de la informalidad muestran estabilidad; la variable migrante sigue siendo uno de los predictores más poderosos, y aunque su coeficiente disminuye en 2024, su permanencia reafirma que el origen nacional es todavía una barrera importante en el mercado laboral local. De la misma manera, los niveles de educación Media y Alta se comportan como factores protectores con coeficientes negativos, mientras que las ocupaciones como Independiente o Precario muestran los coeficientes positivos más altos, confirmando que la informalidad está intrínsecamente ligada a la forma en que los individuos se vinculan al mercado. Es importante señalar que los coeficientes del modelo logit mostrados en las tablas no deben interpretarse directamente como cambios porcentuales en la probabilidad de ser informal, debido a la naturaleza no lineal de la función logística y al uso de log-odds. Por ello, y con el fin de que los resultados sean más fácilmente comprensibles y tengan una lectura económica clara, en la sección siguiente se presentará el análisis de las brechas de informalidad a partir de los Efectos Marginales Promedio (AME), ya que este indicador permite transformar los resultados en cambios directos de puntos porcentuales sobre la probabilidad de ser informal.

Tabla 2. *Presentación de coeficientes de variables y significancia modelo 2024.*

Variable	Coeficiente	Error estándar	Valor t	p-Valor
(Intercepto)	0.749	0.216	3.465	0.0005
migranteMigrante	1.512	0.156	9.656	< 2e-16
edad	-0.021	0.003	-5.334	9.88e-08
sexoMujer	0.686	0.102	6.710	2.09e-11

Variable	Coeficiente	Error estándar	Valor t	p-Valor
educ_levelMedia	-0.934	0.128	-7.295	3.28e-13
educ_levelAlta	-2.219	0.180	-12.303	< 2e-16
ocupacionIndependiente	3.990	0.111	35.674	< 2e-16
ocupacionPrecario/Otros	3.270	0.237	13.752	< 2e-16
estrato_socEstrato Medio-Alto	-0.491	0.150	-3.266	0.001096

Tabla 3. *Presentación de coeficientes de variables y significancia modelo 2021.*

Variable	Coeficiente	Error estándar	Valor t	p-Valor
(Intercepto)	1.377	0,194	7,074	1.64e-12
migranteMigrante	2.500	0,211	11,842	< 2e-16
edad	-0.030	0,003	-8,702	< 2e-16
sexoMujer	0,509	0,094	5,397	6.99e-08
educ_levelMedia	-1,162	0,127	-9,123	< 2e-16
educ_levelAlta	-2,255	0,140	-16,002	< 2e-16
ocupacionIndependiente	4,099	0,116	35,217	< 2e-16
ocupacionPrecario/Otros	3.800	0,221	17,135	< 2e-16
estrato_socEstrato Medio-Alto	-0,748	0,149	-4,998	5.93e-07

4.3 Efectos Marginales Promedio

Con el fin de facilitar la interpretación de los resultados del modelo logit, se estimaron los Efectos Marginales Promedio (AME) para cada una de las variables explicativas. Esta forma de trabajar permite expresar los resultados como cambios en la probabilidad de ser informal, medidos en puntos porcentuales, con lo que se puede entender claramente cuál es el peso relativo de cada factor sobre la informalidad laboral.

Los AME permiten observar de forma más sencilla que los coeficientes originales del modelo logit, la variación promedio de la probabilidad de informalidad ante cambios en las características individuales, manteniendo constantes las demás variables. Sin embargo, como los

AME proporcionan una medida del efecto promedio en un punto, se complementa el análisis con gráficos de las probabilidades predichas.

El motivo de incluir estas gráficas es que, a diferencia de los AME, los valores de probabilidad predicha permiten visualizar interacciones no lineales y ver de forma intuitiva cómo se comportan las brechas de informalidad entre diferentes grupos (como migrantes frente a no migrantes) a lo largo de los años. Los AME nos dan una estimación de impacto promedio, mientras que las probabilidades predichas nos permiten ver cambios en las dinámicas estructurales y focos de vulnerabilidad que no siempre se capturan con el efecto marginal lineal.

Estos resultados permiten identificar con precisión los determinantes de la informalidad en Norte de Santander, y observar las variaciones que se presentan entre 2021 y 2024. En los siguientes puntos, se desarrolla el análisis detallado de los efectos marginales acompañado de las gráficas de probabilidades predichas que facilitan la interpretación de los patrones de comportamiento laboral identificados para el departamento.

Figura 1. Resultados AME año 2021.

	Term	Contrast	Estimate	Std. Error	z
edad	dY/dX		-0.00194	0.000224	-8.69
educ_level	Alta - Baja		-0.17025	0.012230	-13.92
educ_level	Media - Baja		-0.07652	0.008370	-9.14
estrato_soc	Estrato Medio-Alto - Estrato Bajo		-0.05063	0.010654	-4.75
migrante	Migrante - No migrante		0.13675	0.008006	17.08
ocupacion	Independiente - Asalariado		0.52991	0.013745	38.55
ocupacion	Precario/Otros - Asalariado		0.51832	0.015817	32.77
sexo	Mujer - Hombre		0.03229	0.005893	5.48

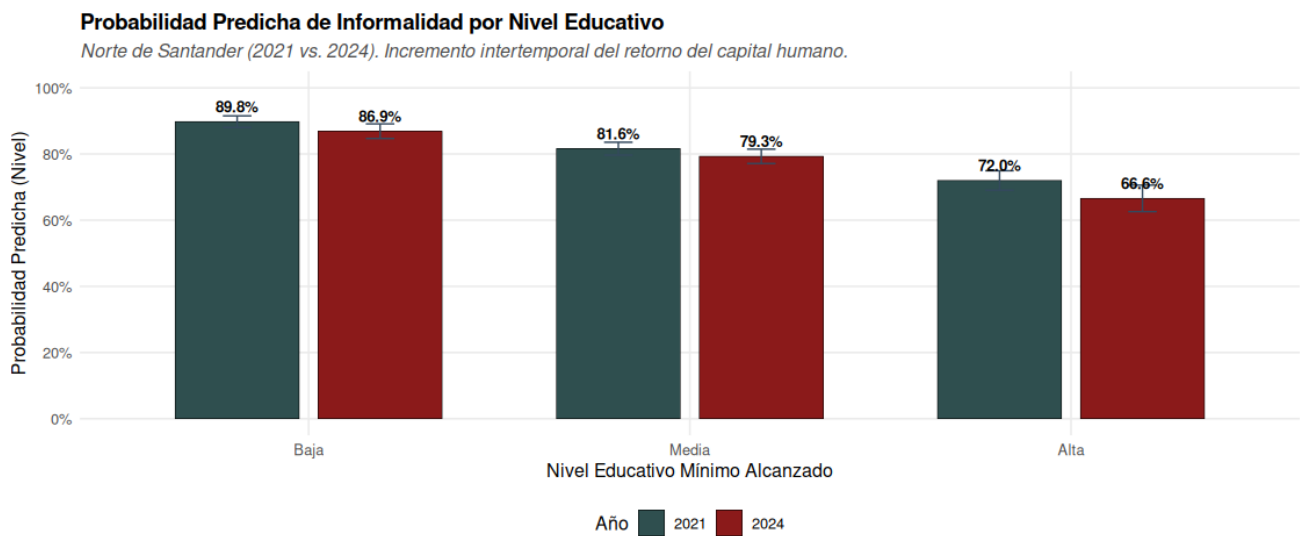
Figura 2. Resultados AME año 2024.

	Term	Contrast Estimate	Std. Error	z
edad	dY/dX	-0.0016	0.000298	-5.38
educ_level	Alta - Baja	-0.2008	0.018201	-11.03
educ_level	Media - Baja	-0.0746	0.010060	-7.42
estrato_soc	Estrato Medio-Alto - Estrato Bajo	-0.0380	0.011988	-3.17
migrante	Migrante - No migrante	0.1088	0.010101	10.77
ocupacion	Independiente - Asalariado	0.5706	0.014534	39.26
ocupacion	Precario/Otros - Asalariado	0.5306	0.021528	24.65
sexo	Mujer - Hombre	0.0508	0.007276	6.99

4.3.1 Efecto del Nivel Educativo sobre la Probabilidad de Informalidad

El capital humano se ratifica como uno de los factores más sólidos que determinan la formalización laboral en Norte de Santander, donde se puede observar en la figura 1 una relación inversa evidente entre el nivel educativo y la probabilidad de informalidad, tendencia que se mantiene estable tanto en 2021 como en 2024.

Figura 3. Gráfico Probabilidad de informalidad según el nivel educativo.



En el transcurso de los dos años se aprecia una disminución progresiva de esta probabilidad: mientras que para los trabajadores con nivel educativo bajo la informalidad supera el 86% en ambos periodos (89,8% en 2021 y 86,9% en 2024), esta cifra desciende de manera notable para quienes alcanzan el nivel medio (81,6% en 2021 y 79,3% en 2024) y se reduce considerablemente en el nivel alto (72,0% y 66,6% respectivamente). Es importante resaltar que estos resultados visuales coinciden totalmente con los Efectos Marginales Promedio (AME) estimados, que cuantifican que comparado con el nivel educativo bajo, tener una educación media reduce la probabilidad de informalidad en 7,6 puntos porcentuales (p.p.) en 2021 y 7,4 p.p. en 2024, mientras que tener una educación alta reduce dicha probabilidad en 17,025 p.p. para 2021 y 20,08 p.p. para 2024.

Esta convergencia entre la representación gráfica y los AME confirma que, si bien en todos los segmentos disminuyó la probabilidad de informalidad, la brecha protectora otorgada por la educación superior se ha fortalecido intertemporalmente, lo que implica un incremento del retorno del capital humano en el mercado regional. Si bien se ha dado esta mejora, la persistencia de una probabilidad superior al 66% incluso en niveles educativos altos refleja que la educación es una condición necesaria pero no suficiente, por sí sola, para garantizar la formalidad en un entorno caracterizado por la baja productividad, subrayando que para los segmentos de menor escolaridad las barreras estructurales de entrada a la formalidad en el departamento son extremadamente rígidas, limitando las oportunidades de una vinculación laboral protegida por la seguridad social.

4.3.2 Efecto de la Migración sobre la Probabilidad de Informalidad

Los resultados demuestran que en los años 2021 y 2024 la condición migratoria se relaciona de forma significativa y sostenida con mayores probabilidades de informalidad laboral

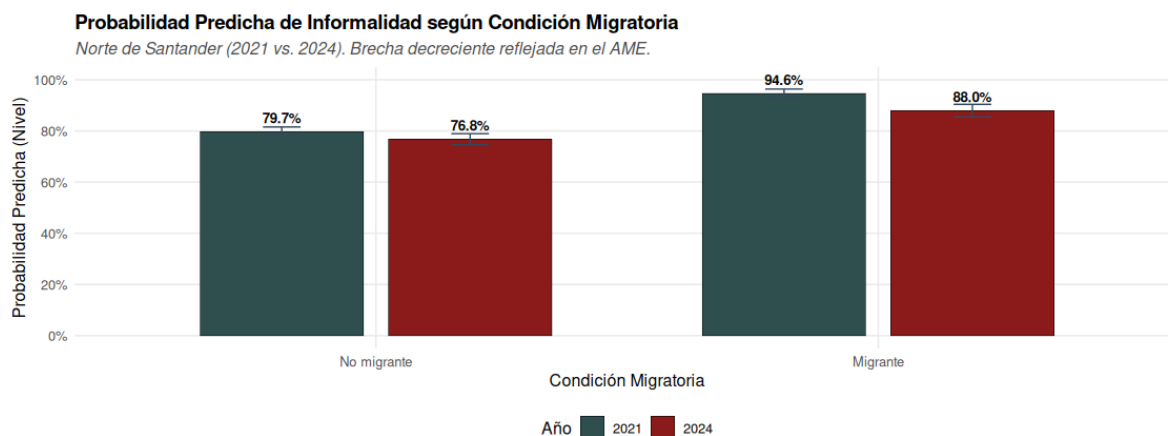
en el departamento de Norte de Santander, lo cual refleja las complejas barreras estructurales a las que esta población se enfrenta para acceder a ocupaciones formales. Los valores estimados de los AME ratifican esta condición, ya que el hecho de ser migrante elevó la probabilidad de informalidad en 13,68 puntos porcentuales (p.p.) en 2021, cifra que se redujo hacia 2024, situándose en 10,88 p.p., lo que apunta a una evolución gradual en los procesos de inserción laboral de los trabajadores extranjeros en la región.

Este fenómeno puede observarse en las probabilidades estimadas que se presentan en la figura 2, donde puede notarse que la condición de migrante aumenta significativamente la probabilidad de informalidad respecto a la población local tanto para el año 2021 como para el 2024. Pero se debe destacar una importante limitación técnica del análisis realizado: la información de la Gran Encuesta Integrada de Hogares (GEIH) no permite captar datos de permisos laborales, ni de los procesos de regularización migratoria, por lo tanto, los resultados no tienen en cuenta el impacto de las políticas de legalización vigentes. Esta limitación, sin embargo, la continuidad de niveles de informalidad notablemente superiores en la población migrante pone en notoriedad una vulnerabilidad de tal profundidad, que rebasa la coyuntura. Diversos factores restrictivos explican esta situación, tales como obstáculos al reconocimiento de títulos académicos o de experiencia laboral técnica, difícil acceso a las redes tradicionales de empleo formal, y alta vulnerabilidad económica con que llega mucha de esta gente al país obligándolos a aceptar empleos de baja productividad o en condiciones precarias.

En el caso particular de Norte de Santander, esa condición fronteriza hace que la situación sea más que crítica. Los flujos migratorios continuos están generando una presión estructural en el mercado laboral regional, intensificando la competencia, sobre todo en los sectores de peor calidad. En conclusión, si bien los resultados entre 2021 y 2024 muestran una lenta mejoría en la

inserción laboral, la informalidad sigue siendo una característica estructural que define a este grupo poblacional.

Figura 4. Probabilidad de informalidad: Migrante VS Local



4.3.3. Efecto de la Edad sobre la Probabilidad de Informalidad

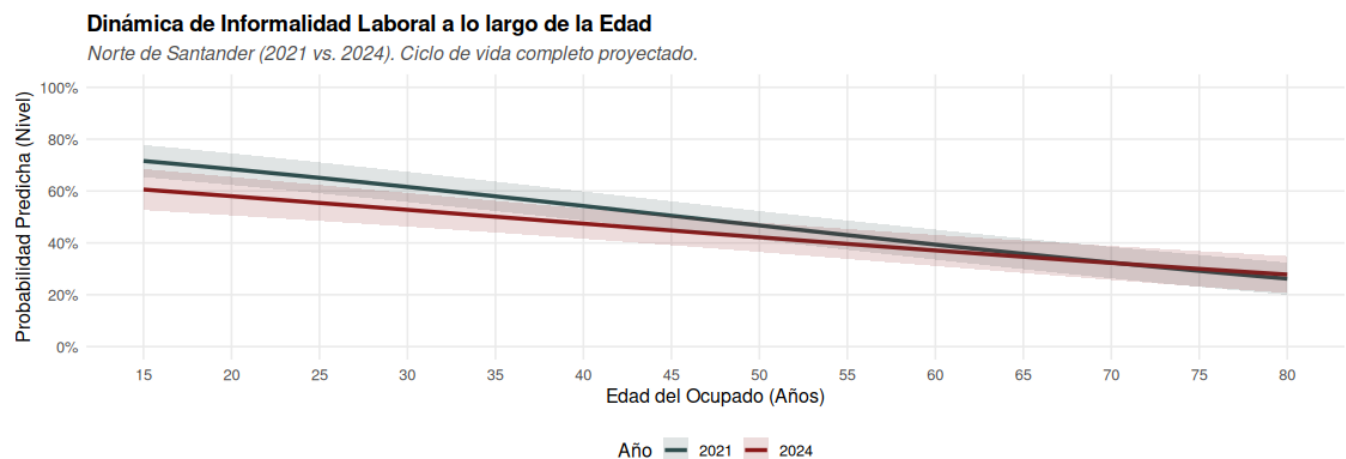
Los resultados sobre la dinámica de la informalidad a lo largo del ciclo de vida, indican que hay una relación negativa estadísticamente significativa entre la edad del ocupado y la probabilidad de estar en informalidad en Norte de Santander. Los AME estimados para el modelo indican que, cada año adicional de edad del trabajador, reduce en 0,19 puntos porcentuales (p.p.) la probabilidad de informalidad en 2021 y en 0,16 p.p. en 2024. Esa tendencia muestra que a medida que los trabajadores acumulan experiencia y madurez laboral van saliendo de la precariedad y pueden ir transitando de forma gradual a ocupaciones con mejores estándares de protección social.

En la figura 3 se puede observar claramente esta relación, las curvas descendientes de probabilidad muestran como los trabajadores más jóvenes enfrentan una mayor exposición a la informalidad al inicio de su vida productiva, brecha que se reduce de forma consistente conforme avanza la edad en ambos periodos analizados. Esto explica el proceso de acumulación de capital

humano específico, de adquisición de experiencia relevante y de acceso progresivo a redes de empleo más estables que actúan como mecanismos protectores frente a las dinámicas informales del mercado regional.

Esta mejora de la estructura del mercado, confirma el modelo, no ha conseguido eliminar la informalidad, que persiste de forma mayoritaria en los primeros escalones de la trayectoria laboral donde la ausencia de una trayectoria consolidada y la alta vulnerabilidad inicial configuran una barrera de entrada persistente al sector formal. En conclusión, los datos confirman que el factor edad tiene un papel determinante en la formalización, siendo la madurez laboral un elemento clave para reducir la brecha de informalidad, sin embargo, el hecho de que la pendiente sea persistente nos sugiere que los jóvenes siguen siendo el segmento más desprotegido.

Figura 5. Probabilidad de informalidad según la edad

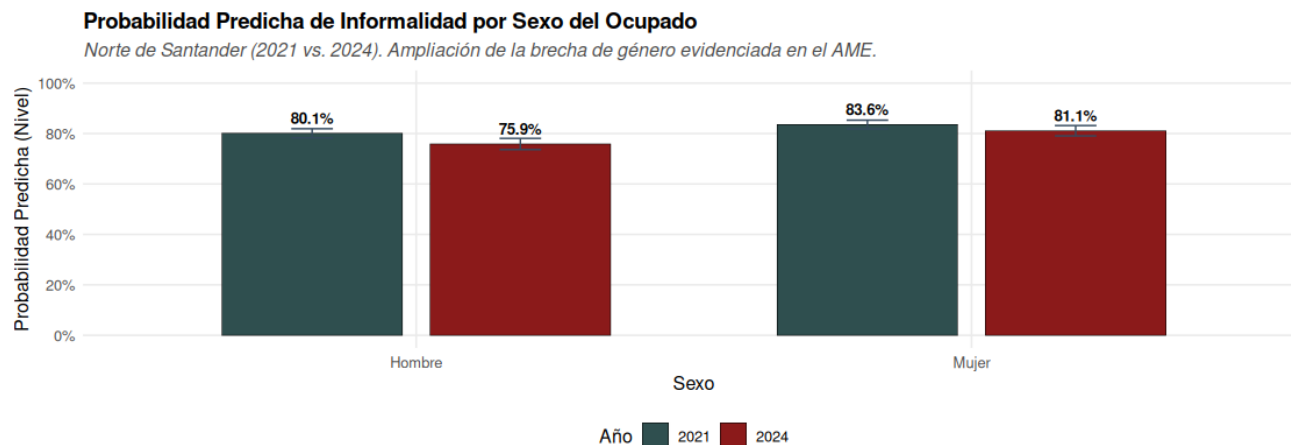


4.3.4 Diferencias por Sexo en la Probabilidad de Informalidad

Los resultados asociados al factor género hacen evidente la existencia de una brecha persistente en la inserción laboral formal dentro del departamento de Norte de Santander, donde las mujeres

se encuentran expuestas a una mayor probabilidad de informalidad frente a sus pares masculinos. Los Efectos Marginales Promedio (AME) señalan que ser mujer eleva la probabilidad de informalidad en 3,23 puntos porcentuales (p.p.) en 2021, valor que crece significativamente a 5,08 p.p. para el año 2024. Ese aumento del AME sugiere que, lejos de cerrarse, la brecha de género en el acceso a empleos con protección por la seguridad social se ha profundizado en el período analizado, lo que implica que los avances generales en la formalización regional no han favorecido a las mujeres en la misma medida que a los hombres.

Esto se ilustra bien en la figura 4, donde se puede observar que la probabilidad de informalidad predicha para las mujeres siempre está por encima de la de los hombres en ambos años. Aunque se observe una ligera disminución en las probabilidades absolutas tanto de hombres como de mujeres entre 2021 y 2024, que esta brecha no se elimine indica que hay barreras estructurales que afectan desproporcionadamente a la mujer en el mercado laboral regional. Esta brecha generalmente está asociada a factores socioeconómicos estructurales, tales como la sobrecarga desequilibrada de las labores de cuidado no remuneradas, que limitan la disponibilidad de tiempo para empleos formales; la segregación ocupacional en sectores de menor productividad donde las mujeres tienen mayor participación y mayores dificultades para acceder a redes de empleo formal que faciliten una vinculación protegida. Los datos ratifican que el género es un eje fundamental de desigualdad en la estructura del mercado laboral nortesantandereano.

Figura 6. *Evolución de la brecha de informalidad por la variable "Sexo"*

4.3.5 Efecto del Estrato Socioeconómico sobre la Probabilidad de Informalidad

Los resultados de la evaluación de impacto del estrato socioeconómico sobre la probabilidad de informalidad laboral para Norte de Santander, confirman que el nivel de ingreso del hogar incide de forma significativa sobre la calidad del empleo. Con efectos al presente análisis se construyó la variable *estrato_soc* mediante la aglutinación de los estratos reportados en la encuesta original en dos grandes categorías: Estrato Bajo (en el cual se encuentran los estratos 1 y 2) y Estrato Medio-Alto (en el cual se encuentran los estratos 3 al 6). Se adoptó esta clasificación con el objetivo de capturar de forma más robusta las disparidades de la estructura socioeconómica regional y mejorar la eficiencia del modelo logístico evitando categorías con representación insuficiente.

Los Efectos Marginales Promedio (AME) muestran que pertenecer a un estrato medio-alto en comparación con el estrato bajo reduce la probabilidad de estar en la informalidad en 5,06 puntos porcentuales (p.p.) en 2021 y en 3,80 p.p. en 2024. Esta relación confirma que, aunque el peso de esta variable ha mostrado una ligera disminución intertemporal, el estrato socioeconómico sigue siendo un predictor importante de la inserción laboral formal del departamento.

Esta dinámica refleja que los hogares que poseen mayores niveles de ingreso socioeconómico, por lo general, cuentan con redes de capital social más amplias y con más recursos para acceder a trayectorias educativas o laborales que facilitan la entrada a sectores con mejores condiciones de protección social. Por el contrario, la población del estrato bajo (la categoría base del modelo) está más expuesta a la informalidad, lo que consolida un ciclo de vulnerabilidad donde la precariedad laboral y las carencias económicas se alimentan entre sí. El efecto persistente indica que los obstáculos estructurales a la inserción en la formalidad no son homogéneos, sino que dependen del origen socioeconómico de los individuos, restringiendo así la movilidad social en el mercado laboral regional.

En conclusión, los resultados ratifican que el estrato socioeconómico, es un factor determinante para diferenciar las oportunidades de acceso al empleo protegido en Norte de Santander. La caída del nivel del AME para el año 2024 sugiere cambios en la composición del mercado que estarían amortiguando ligeramente la influencia directa del estrato, aunque esta variable sigue siendo un componente esencial para explicar la desigualdad en la calidad del trabajo. Por ello, las estrategias de formalización deben tener en cuenta este gradiente socioeconómico, a fin de disminuir las brechas existentes y generar condiciones equitativas de inserción para todos los grupos de la población, independientemente de su estrato de procedencia.

4.3.6 Efecto de la Ocupación sobre la Probabilidad de Informalidad

La posición ocupacional se muestra como una variable clave para describir la calidad del empleo. La variable para este análisis se construyó a partir de la clasificación de posición ocupacional reportada en la GEIH (variable P6430), la cual, originalmente, detalla diferentes tipos de vinculación laboral. Con el fin de agrupar el mercado en función de la naturaleza de la relación,

se procedió a recodificar la información en tres grupos analíticos: “Salario” (que incluye a los obreros y empleados del sector público y privado), “Autónomo” (que incluye a los trabajadores por cuenta propia, patrones o empleadores) y “Precario/otros” (que incluye a los trabajadores familiares no remunerados y otras formas de ocupación sin vínculo formal). Esta clasificación permite ubicar al “Asalariado” como categoría de referencia, ya que bajo la tipología oficial de la encuesta es el grupo que mantiene una relación de subordinación laboral, de modo que facilita la comparación de cómo estas distintas modalidades inciden en la probabilidad de acceso a la protección social.

La evaluación de los resultados obtenidos mediante el modelo logístico permite observar que existe una marcada brecha en la probabilidad de informalidad según la categoría ocupacional del trabajador en Norte de Santander. Los trabajadores con categoría “Independiente”, frente a la categoría de referencia “Asalariado”, presentan un aumento de probabilidad de estar en la informalidad de 52,99 puntos porcentuales (p.p.) en 2021 y de 57,06 p.p. en 2024, de acuerdo con los Efectos Marginales Promedio (AME). La categoría “Precario/Otros”, por su parte, eleva la probabilidad en 51,83 p.p. en 2021 y 53,06 p.p. en 2024, respecto al mismo grupo base.

Estos resultados, que siguen siendo estadísticamente significativos en los dos periodos estudiados, evidencian una marcada segmentación en el mercado laboral regional. El sector asalariado es el que tiene más garantías de formalidad, mientras que los trabajadores independientes y los que están en el grupo de precario o “otros”, tienen mayor exposición a la informalidad. Esa diferencia confirma que la posición ocupacional que depende del tipo de encuesta es el factor más importante para la exclusión de los beneficios de seguridad social. La magnitud de las brechas observadas confirma que las condiciones de contratación asociadas a cada posición ocupacional son el motor fundamental de la precariedad laboral, lo que evidencia una

estructura de mercado donde la ausencia de una vinculación asalariada formal condiciona severamente la calidad del empleo en Norte de Santander.

4.3.7 Síntesis de los resultados

El AME para el periodo 2021-2024, muestra que para el caso de Norte de Santander se trata de una estructura de informalidad persistente y altamente segmentada. Si se comparan ambos años se observa que las variables mantienen una gran estabilidad en su capacidad explicativa, lo que confirma que los determinantes de la informalidad no son fluctuaciones coyunturales, sino elementos estructurales arraigados en el mercado laboral regional. El mayor predictor es la ocupación: el trabajador independiente o con ocupación precaria tiene más de 50 puntos porcentuales (p.p.) de probabilidad de informalidad en comparación con los asalariados, incluso con un aumento de la brecha para los independientes, que pasó de 52,99 p.p. en 2021 a 57,06 p.p. en 2024.

En cuanto a las variables individuales, el nivel educativo tiene un efecto protector robusto, de modo que el nivel “Alta” reduce la probabilidad de informalidad en 17,03 p.p. en 2021 y hasta 20,08 p.p. en 2024 respecto al nivel “Baja”. La edad tiene una relación inversa consistente, aunque de efecto marginal, cada año adicional de vida reduce levemente la probabilidad de la informalidad. Por otro lado, la condición de migrante sigue siendo un factor de riesgo crítico, con un impacto que, si bien se ajustó de 13,68 p.p. a 10,88 p.p., sigue manteniendo una significancia estadística muy robusta. Las disparidades de género también se han ampliado, ya que la brecha de informalidad para las mujeres respecto de los hombres pasó de 3,23 p.p. a 5,08 p.p. en el último año. Por último, el estrato socioeconómico (“Medio-Alto”) continúa funcionando como mecanismo de menor probabilidad de informalidad respecto al “Estrato Bajo”, lo cual confirma

que la capacidad contributiva del hogar es un filtro persistente en la calidad del empleo. Estos resultados exhiben, en conjunto, un mercado laboral regional rígido estructuralmente, donde las condiciones de contratación y el perfil socioeconómico del trabajador actúan como determinantes determinantes sistemáticos que perpetúan la exclusión de la protección social.

5. Prueba de Wald: validación global del modelo

Con el propósito de evaluar la significancia de las variables explicativas consideradas en el modelo se aplicó la prueba de Wald para los años 2021 y 2024. Esta herramienta facilita el análisis econométrico al permitir contrastar hipótesis sobre uno o varios coeficientes de manera simultánea, determinando si Las variables independientes permiten explicar la variable dependiente. (Wooldridge, 2015; Greene, 2012).

En este contexto, Según la hipótesis nula, los coeficientes asociados a las variables explicativas son conjuntamente nulos, es decir:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

H_1 : *Al menos uno de los coeficientes es distinto de cero*

El rechazo de la hipótesis nula en la prueba de Wald global confirma que el modelo propuesto es estadísticamente relevante, lo que valida que el conjunto de variables seleccionadas aporta información significativa para explicar la probabilidad de inserción en la informalidad laboral en Norte de Santander. Los resultados obtenidos muestran que, para el año 2024, el estadístico F de Wald fue de 188,06 con 8 y 7498 grados de libertad, mientras que para 2021 se obtuvo un estadístico F de 204,82 con 8 y 7465 grados de libertad. En ambos periodos, el $P - Valor$ asociado es inferior a 0,001 $p < 2,22 \times 10^{-16}$ lo que permite rechazar la hipótesis nula de nulidad de los coeficientes a cualquier nivel de significancia convencional.

Figura 7. *Resultados de la Prueba Wald en R.*

```

=== TEST DE WALD GLOBAL 2021 ===
> regTermTest(modelo_21, ~ migrante + edad + sexo + educ_level + ocupacion + es
trato_soc)
Wald test for migrante edad sexo educ_level ocupacion estrato_soc
in svyglm(formula = informal ~ migrante + edad + sexo + educ_level +
ocupacion + estrato_soc, design = dis_21, family = quasibinomial())
F = 204.8177 on 8 and 7465 df: p= < 2.22e-16
> cat("\n=== TEST DE WALD GLOBAL 2024 ===\n")

=== TEST DE WALD GLOBAL 2024 ===
> regTermTest(modelo_24, ~ migrante + edad + sexo + educ_level + ocupacion + es
trato_soc)
Wald test for migrante edad sexo educ_level ocupacion estrato_soc
in svyglm(formula = informal ~ migrante + edad + sexo + educ_level +
ocupacion + estrato_soc, design = dis_24, family = quasibinomial())
F = 188.0619 on 8 and 7498 df: p= < 2.22e-16
Session restored from your saved work on 2026-May-31 21:33:03 UTC (5 minutes ag
o),

```

El valor de la significancia estadística global es menor, lo que confirma que variables como migración, edad, sexo, nivel educativo, posición ocupacional y estrato socioeconómico, son importantes para explicar la probabilidad de que una persona esté trabajando en condiciones de informalidad en el departamento. La prueba de Wald comprueba la importancia de todo el conjunto de variables, pero para analizar a fondo el efecto específico de cada variable (tanto en dirección como en magnitud) se hace uso de la interpretación de los coeficientes y los efectos marginales promedios estimados ya analizados.

Los resultados para 2021 y 2024 exhiben un comportamiento altamente consistente, lo que sugiere que los factores determinantes de la informalidad laboral han sido estables en el tiempo, consolidándose incluso después de los ajustes observados en la dinámica del mercado laboral regional en el periodo postpandemia. En su conjunto, la prueba de Wald le da aval estadístico al modelo, lo cual da mayor robustez metodológica a los hallazgos expuestos en el presente estudio, lo cual confirma que la estructura de los determinantes de la informalidad en Norte de Santander no responde a factores aleatorios sino a una configuración socioeconómica persistente.

6. Prueba de robustez

Para garantizar la solidez técnica de los hallazgos y verificar que las conclusiones no dependen exclusivamente de una definición operativa específica, se llevó a cabo una prueba de robustez mediante la reestimación del modelo logístico bajo una métrica alternativa centrada en la protección efectiva en salud. El modelo principal estudió la informalidad bajo el ángulo de la formalidad contractual; esta prueba, en cambio, utiliza la seguridad social como filtro de desprotección real, ya que el acceso a la salud es el pilar del contrato social y el indicador más sensible del bienestar laboral. Para este ejercicio se construyó la variable dependiente “informal_salud” usando exclusivamente las variables P6100 (tipo de seguridad social en salud) y P6110 (quién realiza el pago de la cotización), lo que permitió considerar como informal a todo individuo que reporta estar en el Régimen Subsidiado ($P6100 = 3$) “población sin capacidad de pago contributivo” así como a aquellos trabajadores en el “régimen contributivo” ($P6100 = 1$) que no reportan el pago efectivo de aportes, ya sea por parte del empleador o por cuenta propia. Se hizo una depuración técnica de la base de datos para garantizar la comparabilidad de los sujetos. Se excluyeron las categorías ‘No sabe/No informa’ por ser datos no observables que inducirían sesgo estadístico. Por otra parte, los trabajadores que se encuentran dentro de regímenes especiales (Especial ? (fuerzas armadas, ecopetrol, universidades públicas)) fueron excluidos del análisis porque operan bajo condiciones de protección social y de vinculación laboral que están fuera de la dinámica de mercado del trabajador promedio; debido a que se trata de un grupo con un blindaje institucional distinto, su inclusión habría distorsionado la estimación de los determinantes de la informalidad para el grueso de la población nortesantandereana, impidiendo identificar los riesgos reales de exclusión laboral que buscamos medir. La elección de estos códigos de la GEIH es fundamental, pues permite aislar la desprotección contributiva real, que muchas veces está oculta

tras acuerdos laborales verbales o formas de vinculación indirectas que el modelo principal, por su propia naturaleza, podría clasificar de otra manera.

En comparación de los Efectos Marginales Promedio (AME) obtenidos en este modelo de robustez frente a los del modelo principal, los resultados muestran una consistencia estadística contundente que fortalece la validez de esta investigación. Es importante observar que, aunque las magnitudes absolutas de los efectos en este modelo son menores a las del modelo contractual, esto es esperado y consistente con la teoría económica, dado que la definición basada en la seguridad social es un criterio mucho más estricto y restrictivo; por tanto, la diferencia de escala no invalida los resultados, sino que demuestra que el fenómeno persiste incluso bajo los estándares más exigentes de protección. Si bien es una diferencia de magnitud, la dirección del efecto y la significancia estadística son consistentes en ambos periodos (2021 y 2024). Por ejemplo, la variable de ocupación se reafirma como el motor explicativo más crítico: en 2024, el efecto de ser trabajador independiente o estar en ocupaciones precarias incrementa significativamente la probabilidad de estar en una situación de informalidad contributiva, manteniendo niveles de significancia sólidos ($p < 0.001$). También los factores de control mantienen su rol protector: el nivel educativo muestra que cuanto más se incrementa la escolaridad menor es la probabilidad de desprotección, con un impacto negativo de -0,030 en 2024, mientras que la condición de migrante sigue actuando como un factor de riesgo significativo, con un efecto positivo de 1,58 puntos porcentuales en 2024, replicando el patrón observado en el modelo contractual principal.

Esta estabilidad de los resultados, en la que variables fundamentales como la edad, el sexo y el estrato socioeconómico conservan su comportamiento, permite concluir con rigor que los hallazgos obtenidos no son un artefacto estadístico ni dependen de la arbitrariedad en la elección de la definición operativa. En última instancia, la configuración de la precariedad laboral en Norte

de Santander es una realidad estructural: los grupos poblacionales más vulnerables en términos de capital humano y capacidad contributiva son, consistentemente, los que presentan mayores probabilidades de exclusión, independientemente de si se mide la informalidad por el contrato de trabajo o por la cobertura real en seguridad social. Esta prueba de robustez no solo enriquece la fiabilidad académica del modelo principal, sino que confirma que la informalidad en el departamento es un fenómeno profundo y persistente, cuya naturaleza estructural permanece inmutable frente a diferentes enfoques metodológicos, consolidando la evidencia necesaria para orientar políticas públicas de formalización que vayan más allá de la simple regulación contractual y se centren en la protección integral del trabajador nortesantandereano.

7. Informalidad laboral en el contexto postpandemia (2021–2024)

La razón por la cual el periodo estudiado se centra entre los años 2021 y 2024, es porque en este periodo se evidencian los cambios que se producen en el mercado laboral a medida que va avanzando la recuperación de la crisis de salud ocasionada por el COVID-19, siendo un periodo clave en el análisis de las tendencias del mercado de trabajo en la región Norte de Santander. En 2021 el mercado laboral regional continuaba reflejando los efectos disruptivos de la pandemia con una caída pronunciada del empleo formal y un incremento sustancial de las actividades de subsistencia y el autoempleo, arribando a una tasa de informalidad del 83,17%. Así se puede observar cómo la recuperación inicial de empleo se dirigió mayoritariamente a la informalidad. Los indicadores agregados muestran una ligera mejoría, ya que la informalidad se redujo al 78,88% en 2024, pero este avance debe observarse con cuidado, ya que el modelo econométrico muestra la existencia de brechas estructurales en lugar de una transformación real del tejido productivo.

El análisis econométrico permite concluir que la segmentación del mercado laboral nortesantandereano se mantiene sin grandes cambios, la ocupación se consolida como el factor determinante de la informalidad, lo cual evidencia que la estructura del empleo no ha sufrido mejoras importantes durante la etapa de recuperación. La prueba de robustez basada en la seguridad social confirma esta rigidez, pues muestra que la precariedad laboral prevalece sobre los trabajadores independientes y los empleos menos calificados, sin importar la definición operativa. También la condición migratoria sigue impactando positiva y significativamente sobre la informalidad, lo cual evidencia la existencia de barreras persistentes de acceso que excluyen a esta población de la formalidad, incluso en periodos de crecimiento económico.

Por otra parte, no sólo no se han cerrado las brechas de género, sino que se observa una tendencia de ensanchamiento hacia el 2024. Esto indica que la recuperación del empleo no ha sido equitativa, y afecta desproporcionadamente a las mujeres, que siguen concentradas en sectores informales y asumen una carga mayor de trabajo de cuidados no remunerado. También el nivel educativo y el estrato socioeconómico son variables determinantes, ya que los resultados revelan que los beneficios de la recuperación económica han sido apropiados mayoritariamente por los individuos con mayor escolaridad y capacidad contributiva, dejando a los grupos más vulnerables sin avances significativos.

En conclusión, los resultados muestran que la “recuperación” del mercado laboral en Norte de Santander es parcial y no es suficiente. Los indicadores agregados mejoraron poco, en un contexto de débiles derechos laborales y una informalidad muy extendida. Las causas de riesgo detectadas persisten (confirmadas tanto en el modelo principal como en la prueba de robustez), lo que evidencia que la precariedad laboral no es un fenómeno circunstancial o coyuntural derivado de la pandemia, sino una característica estructural de larga data. Frente a esa realidad, los hallazgos

revisados hacen visible la apremiante necesidad de generar políticas públicas para la formalización que vayan más allá de la mínima regulación contractual, que apunten realmente a mejorar la calidad del empleo y eliminar las restricciones a la entrada a los grupos de la población históricamente excluidas.

8. Recomendaciones de Política Pública

Con base en la evidencia encontrada se propone una estrategia de formalización multidimensional que vaya más allá de la simplificación administrativa y que aborde las causas estructurales de la precariedad laboral en el Norte de Santander. Se propone generar programas de simplificación de trámites para el registro de empresa y cumplimiento de normas laborales como una forma de disminuir las barreras de entrada a la competencia de micro, pequeñas y medianas empresas (MiPymes). Al bajar los costos burocráticos y operativos, se disminuye el umbral de entrada a la formalidad, incentivando a los empleadores a migrar de la informalidad a esquemas de contratación directa, más aún en sectores de alta utilización de mano de obra, el comercio y la agroindustria.

Otra conclusión fundamental es la de lograr un mayor fortalecimiento del capital humano; ello requiere la conexión de la oferta educativa regional a las demandas reales planteadas por el mercado de trabajo; en este sentido, el modelo econométrico utilizado ha puesto de manifiesto que el nivel educativo es uno de los factores determinantes que más protegen frente a la condición de pobreza. Fomentarlos mediante programas de formación técnica y profesional continua que garanticen a las y los trabajadores la adquisición de las competencias requeridas para ingresar a empleos de mayor calidad y reducir la brecha entre la oferta y los requerimientos de las compañías formales. Así mismo se debe diseñar programas de acompañamiento específico que permitan la

integración laboral de la población migrante y la implementación de sistemas de cuidado que permitan la inserción equitativa de las mujeres en el mercado laboral formal, ya que la condición migratoria y las brechas de género son dos elementos que demuestran importantes factores de riesgo para la formalización y actualmente la carga del trabajo no remunerado es una limitación a la participación de las mujeres en el mercado laboral.

Impulsar la competitividad territorial a través del fomento de las cadenas productivas locales sumando incentivos de asistencia técnica y de acceso a crédito para las empresas que adopten prácticas de formalización integral. Con esta estrategia se pretende que el crecimiento económico de la región cree una demanda estable de trabajadores calificados y fomente la estabilidad laboral en el departamento, asegurando que los beneficios de la recuperación económica sean aprovechados por los sectores más vulnerables. En suma, estas recomendaciones no solo pretenden reducir la carga administrativa de las empresas, sino también mitigar las causas estructurales de la precariedad laboral, propiciando una formalización que sea inclusiva, competitiva y perdurable.

9. Conclusiones

La investigación permitió realizar un análisis detallado de los factores sociodemográficos y laborales que inciden en la informalidad del departamento de Norte de Santander en el periodo 2021-2024. A través del uso de microdatos de la Gran Encuesta Integrada de Hogares (GEIH) y la estimación de modelos de regresión logística con diseño muestral complejo, se aportó evidencia empírica relevante para la comprensión de este fenómeno en un contexto regional de alta vulnerabilidad. Los resultados descriptivos señalan que, aunque la tasa de informalidad bajó del 83,17 % en 2021 al 78,88 % en 2024, esta mejora no es suficiente para alterar el carácter estructural

del problema, ya que la caída se explica por una moderada recuperación del mercado laboral tras la crisis sanitaria y no por una transformación real del tejido productivo regional.

Desde un punto de vista explicativo, el modelo econométrico y los efectos marginales promedio (AME) confirman que la informalidad está fuertemente determinada por las características individuales y ocupacionales. La ocupación se convirtió en el principal factor, reafirmando que la estructura productiva regional sigue localizada en trabajos de baja calidad. De igual manera, el nivel educativo ratificó su papel como factor protector, mientras que el estrato socioeconómico, el sexo y la condición migratoria evidenciaron la persistencia de brechas estructurales. Es importante destacar que, aunque la probabilidad de informalidad de la población migrante disminuyó ligeramente en el 2024, este fenómeno requiere de estudios posteriores que permitan determinar si esto se debe a una efectiva integración laboral o a cambios en los patrones de cotización a la seguridad social, incluso para trabajadores con estatus regular.

Para comprobar la solidez del modelo se volvió a estimar con una definición distinta de informalidad basada en la protección real en salud. Al comparar los resultados de esta métrica (que es más estricta y exigente) con los del modelo principal, se puede observar que, si bien los Efectos Marginales Promedio (AME) tienen una magnitud menor, la dirección del efecto y la significancia estadística de los determinantes clave se mantienen sin cambios. La diferencia de escala no es una divergencia sino una ratificación de la solidez de los hallazgos: la precariedad laboral es una realidad estructural en el Norte de Santander que persiste sin importar la definición operativa que se utilice. El análisis muestra pues que una vez controladas las características individuales la recuperación del mercado laboral ha sido heterogénea, manteniendo y, en algunos grupos, profundizando las disparidades estructurales en lugar de mitigarlas.

La prueba de Wald confirmó la importancia de estos hallazgos y la significancia global de los modelos en ambos años. En su conjunto, el estudio contribuye al análisis de la informalidad al proporcionar evidencia empírica robusta, actualizada y contextualizada, facilitando la identificación de los perfiles más vulnerables para el diseño de políticas públicas focalizadas. El estudio aporta académicamente al uso de metodologías econométricas en el tratamiento de encuestas complejas, combinando análisis descriptivo con modelos explicativos y ejercicios de robustez. En última instancia, la evidencia confirma que la informalidad no es un fenómeno coyuntural derivado de la pandemia, sino un obstáculo estructural de larga data para el desarrollo regional, cuya resolución requiere intervenciones integrales que prioricen la calidad del empleo y la remoción de barreras de acceso para los grupos poblacionales históricamente excluidos.

Referencias

- Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. National Bureau of Economic Research. <https://www.nber.org/books-and-chapters/human-capital-theoretical-and-empirical-analysis-special-reference-education-second-edition/front-matter-human-capital-theoretical-and-empirical-analysis-special-reference-education>
- Becker, G. S. (1993). *Human capital: A theoretical and empirical analysis, with special reference to education* (3.^a ed.). University of Chicago Press.
- Bernal S., R. (2009). El mercado laboral informal en Colombia: identificación y caracterización. *Desarrollo Y Sociedad*, 1(63), 145-208. <https://doi.org/10.13043/dys.63.4>
- Cahuc, P., Carcillo, S., & Zylberberg, A. (2014). *Labor economics* (2nd ed.). MIT Press.
- Congreso de la República de Colombia. (1993, 23 de diciembre). *Ley 100 de 1993. Por la cual se crea el sistema de seguridad social integral y se dictan otras disposiciones*. Diario Oficial No. 41.148.
- Congreso de la República de Colombia. (2002, 27 de diciembre). *Ley 789 de 2002. Por la cual se dictan normas para apoyar el empleo y ampliar la protección social y se modifican algunos artículos del Código Sustantivo de Trabajo*. Diario Oficial No. 45.046.
- Congreso de la República de Colombia. (2010, 29 de diciembre). *Ley 429 de 2010. Ley de Formalización y Generación de Empleo*. Diario Oficial No. 47.937.
- De Soto, H. (1986). *El otro sendero: La revolución informal*. Instituto Libertad y Democracia.
- Departamento Administrativo Nacional de Estadística (DANE). (2020). *Gran Encuesta Integrada de Hogares (GEIH): Boletín técnico. Mercado laboral por departamentos*. chrome-

extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.dane.gov.co/files/investigaciones/boletines/ech/ml_depto/Boletin_dep_20.pdf

Departamento Administrativo Nacional de Estadística (DANE). (2021). *Gran Encuesta Integrada de Hogares (GEIH): Boletín técnico. Mercado laboral por departamentos*. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.dane.gov.co/files/investigaciones/boletines/ech/ml_depto/Boletin_dep_21.pdf

Departamento Administrativo Nacional de Estadística (DANE). (2024). *Gran Encuesta Integrada de Hogares (GEIH): Boletín técnico. Mercado laboral por departamentos*. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.dane.gov.co/files/operaciones/GEIH/bol-GEIHDepartamentos-2024.pdf

Departamento Administrativo Nacional de Estadística. (s. f.). *Glosario*. <https://www.dane.gov.co/index.php/servicios-al-ciudadano/tramites/transparencia-y-acceso-a-la-informacion-publica/informacion-de-interes/glosario>

Doeringer, P. B., & Piore, M. J. (1971). *Internal labor markets and manpower analysis*. M. E. Sharpe. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://files.eric.ed.gov/fulltext/ED048457.pdf

G, W. H. (2018). *Econometric analysis* (8th ed.). Pearson. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://api.pageplace.de/preview/DT0400.9781292231150_A39514649/preview-9781292231150_A39514649.pdf

Galvis, L. A. (2012). Informalidad laboral en las áreas urbanas de Colombia (Borradores de Economía, Núm. 696). Banco de la República. chrome-

extension://efaidnbmnnnibpcajpcglclefindmkaj/https://repositorio.banrep.gov.co/server/api/core/bitstreams/94578ae1-1d06-4bd0-91cf-0e3d46d486a9/content

Greene, W. (2012) *Econometric Analysis*. (7th Edition)

Gujarati, D. N., & Porter, D. C. (2010). *Econometría* (5.^a ed.). McGraw-Hill. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://files.uagrm.edu.bo/entidad/161/file/indexed/Ecomod/An%C3%A1lisis%20econom%C3%A9trico%20con%20Eviews%209/Econometria-Damodar-N-Gujarati-5ta%20Ed.pdf

Harris, J. R., & Todaro, M. P. (1970). Migration, unemployment and development: A two-sector analysis. *The American Economic Review*, 60(1), 126–142.

Hart, K. (1973). Informal income opportunities and urban employment in Ghana. *The Journal of Modern African Studies*, 11(1), 61–89. <https://doi.org/10.1017/S0022278X00008089>

Hosmer, D. W., Lemeshow, S., y Sturdivant, R. X. (2013). *Applied Logistic Regression* (3.^a ed.). Wiley.

Lewis, W. A. (1954). Economic development with unlimited supplies of labour. *The Manchester School*, 22(2), 139–191. <https://doi.org/10.1111/j.1467-9957.1954.tb00021.x>

Marshall, A. (2013). *Principles of economics*. (Original work published 1890)

McConnell, C. R., Brue, S. L., & Macpherson, D. A. (2017). *Economía laboral* (11.^a ed.). McGraw-Hill.

Mincer, J. A. (1974). Front matter. En J. A. Mincer, *Schooling, Experience, and Earnings*. National Bureau of Economic Research. <http://www.nber.org/chapters/c1760>

Ocampo, J. A. (Ed.). (2005). *Historia económica de Colombia*. Planeta.

Organización Internacional del Trabajo (OIT). (2002). *El trabajo decente y la economía informal*. Conferencia Internacional del Trabajo, 90.^a reunión.

- Organización Internacional del Trabajo (OIT). (2013). *Resolución sobre las estadísticas del trabajo, la ocupación y la subutilización de la fuerza de trabajo*. 19ª Conferencia Internacional de Estadísticos del Trabajo. <https://www.ilo.org/es/media/267096/download>
- Organización Internacional del Trabajo (OIT). (2015). *Recomendación sobre la transición de la economía informal a la economía formal (núm. 204)*.
- Organización Internacional del Trabajo. (2015, 12 de junio). *Recomendación sobre la transición de la economía informal a la economía formal* (Recomendación núm. 204). <https://share.google/buAcztaetJT9G91P0>
- Organización Internacional del Trabajo. (2018). *Women and men in the informal economy: A statistical picture (3rd ed.)*. International Labour Office.
- Organización Internacional del Trabajo. (2021). *Informe Mundial sobre la Protección Social 2020-22: La protección social en la encrucijada – en busca de un futuro mejor*. <https://www.ilo.org/es/publications/flagship-reports/informe-mundial-sobre-la-proteccion-social-2020-2022-la-proteccion-social>
- Organización Internacional para las Migraciones. (2019). *Informe sobre las migraciones en el mundo 2020*. https://publications.iom.int/system/files/pdf/wmr_2020_es.pdf
- Piore, M. J. (1979). *Birds of passage: Migrant labor and industrial societies*. Cambridge University Press.
- Portes, A., y Haller, W. (2004). *La economía informal* (Serie Políticas Sociales N.º 100). Comisión Económica para América Latina y el Caribe (CEPAL). <https://digitallibrary.un.org/record/544195?v=pdf>
- Portes, A., y Haller, W. (2004). *La economía informal* (Serie Políticas Sociales N.º 100). CEPAL. <https://hdl.handle.net/11362/6091>

Presidencia de la República de Colombia. (2021, 1 de marzo). *Decreto 216 de 2021. Por el cual se adopta el Estatuto Temporal de Protección para Migrantes Venezolanos bajo Régimen de Protección Temporal*. Diario Oficial No. 51.602.

Ramírez Zambrano, J. R., Zambrano Miranda, M. J., Mogrovejo Andrade, J. M., & Carreño Montaña, J. L. (2016). Informalidad laboral en los departamentos de Norte de Santander, Nariño, La Guajira y Cesar. *Apuntes del CENES*, 35(62), 125–145.
<https://doi.org/10.19053/22565779.5232>

Samuelson, P. A., & Nordhaus, W. D. (2010). *Economía* (19.^a ed.). McGraw-Hill.

Schneider, F. (1986). Estimating the size of the Danish shadow economy using the currency demand approach: An attempt. *The Scandinavian Journal of Economics*, 88(4), 643–668.
<https://www.jstor.org/stable/3440435>

Schultz, T. W. (1961). Investment in human capital. *The American Economic Review*, 51(1), 1–17. <https://www.jstor.org/stable/1818907>

Smith, A. (2004). *La riqueza de las naciones*. Alianza Editorial. (Original work published 1776)

Tokman, V. E. (2001). *De la informalidad a la modernidad*. Oficina Internacional del Trabajo.
chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.oitcinterfor.org/sites/default/files/file_articulo/tokman.pdf

UNESCO. (2012). *Clasificación internacional normalizada de la educación: CINE 2011*. Instituto de Estadística de la UNESCO.

Wooldridge, J. M. (2015). *Introducción a la econometría: Un enfoque moderno* (5.^a ed.). Cengage Learning.

Zhou, M. (1997). Growing up American: The challenge confronting immigrant children and children of immigrants. *Annual Review of Sociology*. <https://www.jstor.org/stable/2952544>