

# LECTURAS DE ECONOMÍA

Versión preliminar

**Predicción de la formalidad en las micro y pequeñas empresas del Perú:  
Un modelo basado en factores estructurales y estratégicos**

Raquel Chafloque-Céspedes y Lorena del Carmen Espina-Romero

El presente artículo ha sido aprobado para ser publicado en Lecturas de Economía 104. Sin embargo, se trata de una versión preliminar, la cual está sujeta a cambios asociados al proceso editorial de la revista. Próximamente se contará con una versión definitiva.

Preliminar

# Predicción de la formalidad en las micro y pequeñas empresas del Perú: Un modelo basado en factores estructurales y estratégicos

Raquel Chafloque-Céspedes  
y Lorena del Carmen Espina-Romero<sup>1</sup>

## Resumen

La informalidad empresarial representa un desafío estructural en las economías emergentes, limitando la productividad, el acceso al crédito y la inclusión de las micro y pequeñas empresas (MYPEs) en redes formales. Este artículo desarrolla un modelo predictivo para estimar la probabilidad de formalidad en las MYPEs peruanas a partir de factores estructurales y estratégicos, mediante una regresión logística binaria. Se plantea que características como el tamaño empresarial, la tenencia del local y niveles elevados de producción incrementan significativamente la probabilidad de formalización. Se emplearon datos del V Censo Nacional Económico 2022 ( $n= 175\,264$ ). Los resultados revelan que las empresas del sector comercio y servicios, con mayores niveles de ventas y resultados de ejercicio mayores, incrementan la probabilidad de formalidad, mientras que el tamaño reducido y la operación sin local fijo disminuyen esta probabilidad. En cambio, se encontró que altos niveles de ventas no necesariamente se asocian con procesos de formalización sostenida. Estos hallazgos ofrecen un modelo estadístico útil para focalizar políticas públicas hacia segmentos con mayor riesgo de informalidad. Se recomienda ampliar el análisis con datos longitudinales y variables cualitativas para capturar mejor la complejidad del fenómeno.

**Palabras clave:** formalidad empresarial, micro y pequeñas empresas, modelo predictivo, regresión logística, informalidad en Perú.

**Clasificación JEL:** O17, C25, D22, J46.

## Predicting Formality in Micro and Small Enterprises in Peru: A Model Based on Structural and Strategic Factors

## Abstract

Business informality represents a structural challenge in emerging economies, limiting productivity, access to credit, and the inclusion of micro and small enterprises (MSEs) in formal networks. This study develops a predictive model using binary logistic regression to estimate the probability of formality in Peruvian MSEs based on structural and strategic factors. It is proposed that characteristics such as firm size, tenure of premises and high levels of production significantly increase the probability of formalization. Data from the V

---

<sup>1</sup> *Raquel Chafloque-Céspedes*: profesora de la Universidad San Ignacio de Loyola, Escuela de Posgrado, Lima, Perú. Dirección electrónica: mchafloquec@usil.edu.pe  
<https://orcid.org/0000-0003-0314-5603>

*Lorena del Carmen Espina-Romero*: profesora de la Universidad San Ignacio de Loyola, Escuela de Posgrado, Lima, Perú. Dirección electrónica: [lespina@usil.edu.pe](mailto:lespina@usil.edu.pe)  
<https://orcid.org/0000-0002-6637-8300>

National Economic Census 2022 ( $n= 175\,264$ ) were used. The results reveal that companies in the trade and services sector, with higher levels of sales and higher fiscal year results, increase the probability of formality, while small size and operation without fixed premises decrease this probability. On the other hand, it was found that high levels of sales are not necessarily associated with sustained formalization processes. These findings provide a statistical model that can guide public policies towards segments with a higher risk of informality. It is recommended to expand the analysis with longitudinal data and qualitative variables to better capture the complexity of the phenomenon.

**Keywords:** business formality, micro and small enterprises, predictive modeling, logistic regression, informality in Peru.

### **Prévision de la formalité dans les micro et petites entreprises du Pérou : un modèle basé sur des facteurs structurels et stratégiques**

#### **Résumé**

L'informalité des entreprises représente un défi structurel dans les économies émergentes, limitant la productivité, l'accès au crédit et l'inclusion des micro et petites entreprises (MPE) dans les réseaux formels. Cet article développe un modèle prédictif pour estimer la probabilité de formalisation des MPE péruviennes à partir de facteurs structurels et stratégiques, à l'aide d'une régression logistique binaire. Il suggère que des caractéristiques telles que la taille de l'entreprise, la propriété des locaux et des niveaux de production élevés augmentent considérablement la probabilité de formalisation. Les données utilisées proviennent du V Censo Nacional Económico 2022 ( $n = 175\,264$ ). Les résultats révèlent que les entreprises du secteur du commerce et des services, qui affichent des niveaux de ventes et des résultats d'exploitation plus élevés, augmentent la probabilité de formalisation, tandis que la petite taille et l'absence de locaux fixes diminuent cette probabilité. En revanche, il a été constaté que des niveaux de vente élevés ne sont pas nécessairement associés à des processus de formalisation durable. Ces résultats offrent un modèle statistique utile pour cibler les politiques publiques vers les segments présentant un risque plus élevé d'informalité. Il est recommandé d'élargir l'analyse avec des données longitudinales et des variables qualitatives afin de mieux saisir la complexité du phénomène.

**Mots clés:** formalité entrepreneuriale, micro et petites entreprises, modèle prédictif, régression logistique, informalité au Pérou.

<https://doi.org/10.17533/udea.le.n104a361071>

**Cómo citar / How to cite this item:** Chafloque-Céspedes, R., & Espina-Romero, L. C. (2025). Predicción de la formalidad en las micro y pequeñas empresas del Perú: Un modelo basado en factores estructurales y estratégicos. *Lecturas de Economía*, (104). <https://doi.org/10.17533/udea.le.n104a361071>

*Primera versión recibida el 4 de junio de 2025; versión final aceptada el 4 de agosto de 2025*

**Introducción – I. Revisión de literatura. – II. Metodología y datos. – III. Resultados. –IV. Discusión. –Conclusiones. – Declaración de ética. – Anexo – Referencias.**

### Introducción

La informalidad empresarial continúa siendo uno de los desafíos estructurales más persistentes en las economías en desarrollo, especialmente en América Latina. En países como Perú, el sector informal representa una porción significativa del aparato productivo, albergando a millones de micro y pequeñas empresas (MYPEs) que operan al margen de las regulaciones tributarias, laborales y municipales. Aunque constituyen una fuente importante de empleo e ingresos, enfrentan limitaciones para acceder a servicios financieros, redes comerciales formales y programas estatales de apoyo (de Soto, 1986; La Porta & Shleifer, 2014).

La formalización empresarial ha sido estudiada desde múltiples enfoques teóricos. Por ejemplo, la perspectiva estructuralista resalta la relación entre la informalidad, la baja productividad y las restricciones al crecimiento empresarial, mientras que la teoría institucional enfatiza la capacidad del Estado para generar incentivos efectivos para el cumplimiento normativo (Loayza, 1997). Por su parte, los enfoques de comportamiento racional plantean que la informalidad puede responder a una decisión estratégica de los emprendedores frente a costos regulatorios percibidos como excesivos (Robles & Ambriz, 2023; Tokman, 2001).

Dentro de este campo, un área creciente de investigación busca identificar los factores que influyen en la decisión de formalizarse. Estudios recientes han señalado el papel de elementos estructurales, como el tamaño de la empresa, la tenencia del local y el sector económico, así como la organización jurídica como factor relevante (Günther & Launov, 2012; World Bank, 2020). Por otro lado, factores estratégicos como el volumen de ventas, el nivel de producción o la inversión en capital humano también inciden en la probabilidad de formalización (Aberra et al., 2022; Kede et al., 2023). Sin embargo, a pesar del creciente cuerpo de literatura, persisten importantes vacíos empíricos respecto a cómo estos factores interactúan en el contexto específico de las MYPEs peruanas.

Uno de los principales problemas de los estudios existentes es el uso de muestras reducidas o no representativas, lo que limita la posibilidad de generalizar los resultados. También, que pocos trabajos han utilizado bases censales para modelar la formalidad desde un enfoque predictivo multivariable. En este sentido, aún se carece de modelos estadísticos robustos que permitan estimar con precisión la probabilidad de formalización empresarial, incorporando simultáneamente factores estructurales y estratégicos y contextualizados en el Perú.

Este artículo contribuye a cerrar dicho vacío mediante el desarrollo de un modelo predictivo que estima la probabilidad de formalidad en las MYPEs peruanas. Para ello, se emplearon datos del V Censo Nacional Económico 2022, el cual ofrece información detallada y actualizada sobre las características de las unidades productivas del país. El objetivo central

de este texto es identificar cuáles son los factores que mejor explican la condición de formalidad en este segmento empresarial y, en consecuencia, aportar evidencia empírica que oriente el diseño de políticas públicas más focalizadas.

La metodología aplicada, regresión logística binaria sobre una muestra nacional de 175 264 MYPEs, permite abordar adecuadamente la pregunta de investigación, dado que la variable dependiente es dicotómica (formal vs. informal) y se dispone de múltiples predictores relevantes. Esta técnica, validada ampliamente en estudios sociales y económicos, permite estimar *odds ratios* (razones de momios) para interpretar el peso relativo de cada factor en la probabilidad de formalidad (Hosmer et al., 2013).

El aporte principal de este estudio radica en ofrecer un modelo empíricamente robusto, basado en datos oficiales y de amplia cobertura, que permite comprender con mayor precisión qué perfiles de empresas están en mayor riesgo de informalidad. Asimismo, se formulan recomendaciones dirigidas a la política pública y se proponen futuras líneas de investigación, considerando tanto las limitaciones como las posibilidades de ampliación del modelo propuesto.

Este artículo está organizado en cinco secciones. Tras esta introducción, la primera sección desarrolla el marco teórico, abordando los principales enfoques sobre la informalidad empresarial y los determinantes de la formalidad. La segunda sección describe la metodología utilizada, incluyendo el diseño del estudio, la selección de variables y el enfoque de análisis. En la tercera sección se presenta y analizan los resultados del modelo predictivo. Posteriormente, en la cuarta sección se discuten los hallazgos a la luz de la literatura previa, se proponen implicaciones prácticas, se reconocen las limitaciones del estudio. Finalmente, en las conclusiones se sugieren direcciones para investigaciones futuras.

### I. Revisión de literatura

La informalidad empresarial ha sido abordada desde diversas perspectivas teóricas, entre ellas la estructuralista, la institucional y la del comportamiento racional. En el centro del debate se encuentra la necesidad de comprender las condiciones bajo las cuales las MYPEs deciden operar en la formalidad o en la informalidad. La literatura ha demostrado que la informalidad no es meramente una etapa transitoria hacia la formalización, sino que en muchos contextos constituye una decisión deliberada basada en cálculos económicos y percepciones institucionales (Günther & Launov, 2012; La Porta & Shleifer, 2014).

Una línea importante de investigación se ha centrado en identificar los determinantes de la formalidad o informalidad empresarial. Al respecto, de Soto (1986) argumentó que la complejidad normativa y los altos costos de cumplimiento generan una “muralla legal” que aleja a los pequeños negocios de la formalización. Más recientemente, Loayza (1997) y Tokman (2001) reforzaron esta idea al señalar que la informalidad es un fenómeno estructural asociado a la baja productividad, el acceso limitado a capital y la escasa capacidad institucional del Estado. Ahora, desde una perspectiva estructuralista, el estudio de Aberra et al. (2022) aporta evidencia empírica, destacando que las empresas informales presentan en promedio

menores niveles de educación del emprendedor, menor acceso a financiamiento y débiles vínculos con cadenas de valor formales.

En los últimos años, el análisis empírico sobre la formalidad empresarial ha considerado la aplicación de modelos predictivos como la regresión logística, el aprendizaje automático y las ecuaciones estructurales, los cuales permiten estimar la probabilidad de formalidad empresarial. Por ejemplo, Kede et al. (2023) demostraron que la innovación en productos y procesos aumenta significativamente la probabilidad de formalización y la creación de empleo en pymes informales de Camerún. Asimismo, Mehjabeen y Noor (2023) encontraron, a partir de un modelo logístico, que factores como el deseo de acceso al crédito, el reconocimiento social y la presión institucional son predictivos del tránsito hacia la formalidad en empresas de Bangladesh.

En contextos más amplios, como el mexicano, trabajos recientes han modelado la informalidad con métodos estructurales, incorporando variables macroeconómicas como impuestos, gasto público y desempleo para explicar su evolución de largo plazo (Robles & Ambriz, 2023). Desde una perspectiva microeconómica, diversos estudios han destacado la importancia de factores estructurales como el tamaño, la tenencia del local, la ubicación geográfica, la organización jurídica y el sector económico. Estas características condicionan tanto la capacidad para cumplir con obligaciones legales como su exposición a la fiscalización estatal. Por ejemplo, Buitrago et al. (2024) concluyen que el marco institucional y la calidad del entorno regulatorio influyen más en la prevalencia de la informalidad que el tipo de sector o nivel de ingreso.

Por otro lado, los factores estratégicos se relacionan con decisiones empresariales vinculadas al desempeño, como el nivel de producción, el volumen de ventas y la inversión en capital humano. Estos elementos no determinan directamente la formalización, pero pueden incentivarla o desincentivarla según el contexto institucional. Así, Top y Zhi (2023) encontraron que, en economías emergentes, la orientación emprendedora individual y el capital social influyen significativamente en la expansión de microempresas informales digitales, independientemente de su desempeño económico formal.

En este contexto, los censos económicos y las encuestas nacionales constituyen fuentes clave para modelar la informalidad. A diferencia de encuestas focalizadas, los censos capturan información de unidades productivas registradas y no registradas en múltiples sectores, permitiendo estimar probabilidades de formalización a gran escala. Bravo-Ortega et al. (2024) utilizan estos instrumentos para evidenciar cómo las instituciones informales, como redes sociales y reglas no escritas, pueden facilitar o bloquear la formalización en contextos de incongruencia institucional. En la misma línea, el “World Development Report 2020” destaca el rol de las cadenas globales de valor y el comercio internacional como fuerzas que impulsan la formalización, especialmente cuando existen incentivos claros para integrarse a mercados regulados (World Bank, 2020).

No obstante, este estudio no incluye algunas variables identificadas en investigaciones previas como relevantes para la decisión de formalización, tales como el monto de la inversión inicial, el nivel educativo del emprendedor, el tiempo de dedicación al negocio, las fuentes de

financiamiento, la rentabilidad, las razones de inicio y las formas de pago. Estudios desarrollados en contextos latinoamericanos —como los realizados en México por Alvarado Lagunas (2021) y Rodríguez Medina et al. (2023)— han demostrado que estos factores pueden ejercer una influencia significativa sobre la condición de formalidad, especialmente en el caso de nanoempresas o unidades de subsistencia ubicadas en zonas perimetropolitanas. La omisión de estas variables responde tanto a la disponibilidad de datos en el Censo Nacional Económico (CENE) del 2022 como al enfoque estructural y estratégico adoptado.

Ahora bien, la relación entre volumen de ventas y formalidad empresarial sigue siendo objeto de debate. Algunos estudios sostienen que mayores ventas incrementan la necesidad de formalizarse, especialmente para acceder a financiamiento, expandirse comercialmente o participar en licitaciones públicas (Kede et al., 2023; Mehjabeen & Noor, 2023). Por el contrario, otras investigaciones documentan que ciertos segmentos de MYPEs con altos ingresos optan por permanecer en la informalidad, aprovechando la baja fiscalización o buscando evitar los costos asociados al cumplimiento normativo (de Soto, 1986; Robles & Ambriz, 2024). En el caso peruano, este hallazgo contraintuitivo, que vincula ventas elevadas vinculadas con informalidad, podría explicarse por la racionalidad estratégica de los empresarios en contextos institucionales frágiles, donde la evasión se percibe como un recurso para sobrevivir o maximizar utilidades.

En suma, el marco teórico que sustenta este estudio integra enfoques estructurales e institucionales, metodologías predictivas recientes y evidencia empírica basada en fuentes censales. Esta combinación permite analizar de forma rigurosa las condiciones bajo las cuales una MYPE decide operar formal o informalmente, entendiendo la formalización no solo como un estado legal, sino como un proceso condicionado por múltiples dimensiones internas y contextuales. A partir de la literatura revisada, se justifica el uso de un enfoque cuantitativo para identificar los factores determinantes de la formalidad en las MYPEs.

En consonancia con lo anterior, este estudio se guía principalmente por el enfoque institucional, considerando que la formalización empresarial depende de la capacidad del entorno regulatorio para inducir el cumplimiento normativo. La selección de variables estructurales como la organización jurídica, el tamaño de la empresa y la tenencia del local refleja esta orientación, pues representan mecanismos mediante los cuales las instituciones influyen en las decisiones empresariales. No obstante, se incorporan elementos del enfoque de comportamiento racional, especialmente al analizar casos en los que altos niveles de ventas no derivan en formalización, lo que podría reflejar estrategias deliberadas de evasión.

En ese sentido, el modelo empírico propuesto articula ambos enfoques, con predominancia institucional. Esta elección resulta coherente con el objetivo del estudio, la naturaleza de los datos censales y los resultados obtenidos. A continuación, se describe la metodología empleada para desarrollar y validar el modelo predictivo propuesto.

## II. Metodología y datos

Este estudio adopta un enfoque cuantitativo, con diseño transversal y alcance correlacional-predictivo, el cual permite analizar asociaciones entre variables sin manipulación directa. Esta metodología es ampliamente utilizada en estudios sobre comportamiento organizacional y empresarial (Hernández et al., 2014). Se empleó la regresión logística binaria, adecuada para modelar una variable dependiente dicotómica (formal vs. informal) y estimar efectos marginales a través de *odds ratios*.

#### **A. Selección de la muestra**

La fuente de información primaria fue el V Censo Nacional Económico (CENEC) del 2022, elaborado por el Instituto Nacional de Estadística e Informática (INEI, 2024). Este proporciona una base de datos exhaustiva sobre las unidades empresariales activas en el país, considerando la información económica, financiera y la estructura empresarial. La selección de la muestra se realizó a partir de los registros de MYPEs, definidas por el número de trabajadores y volumen de ventas anuales, conforme a la normativa peruana vigente. El CENEC 2022 cuenta con un total de 1 901 309 empresas ubicadas en el área urbana del país, sin distinción sectorial.

Con el fin de delimitar la población objetivo a MYPEs, se aplicaron los lineamientos de la clasificación oficial de la Superintendencia Nacional de Administración Tributaria (SUNAT, 2025), que indica que el valor de volumen de ventas anuales debe ascender hasta 1700UIT (equivalente a S/.7 480 000 para el año 2021, y el máximo de trabajadores es de 100 personas por empresa. Para realizar la selección de la muestra y la preparación de los datos, se siguió un proceso riguroso. En una primera etapa, se eliminaron todas las empresas que no reportaron ventas en 2021, reduciendo el universo a 1 400 326 casos.

Posteriormente, se filtraron las empresas que efectivamente operaron durante el año de referencia<sup>2</sup>, quedando 764 639 empresas. Al aplicar el tope de ventas y número de trabajadores, se obtuvieron 732 769 registros que cumplían ambos criterios. Luego, se procedió a excluir los casos con datos faltantes en las variables clave (ventas, producción y remuneración) quedando 695 348 empresas. Asimismo, se conservaron únicamente los establecimientos que indicaron con certeza si tenían o no RUC, descartando aquellos que indicaron no saberlo, quedando 622 882 casos.

Como paso siguiente, se excluyeron las empresas que reportaban ventas, pero no producción (o viceversa) de forma contradictoria; por ejemplo, casos donde la producción era igual al valor de ventas sin haber declarado producción explícita. Esto se hizo para evitar duplicidad conceptual entre variables, se obtuvieron 457 633 registros. Dado que el análisis utiliza transformadores logarítmicos<sup>3</sup>, se eliminaron todos los casos con valores de cero o negativos en las variables de ventas, producción o remuneración, por lo que la muestra se redujo a 252 231 casos.

Finalmente, se excluyeron las empresas con resultados negativos en la variable “valor agregado” o “resultado del ejercicio”, ya que estos podrían comprometer la validez

<sup>2</sup> Variable C2P12A\_2021: situación de la empresa.

<sup>3</sup> Se utilizaron logaritmos naturales con el fin de reducir la asimetría y dispersión de los datos, normalizar la escala y facilitar la interpretación.

económica del modelo y distorsionar la interpretación de los predictores. La muestra final quedó conformada por 175 264 MYPEs.

La limpieza y preparación de la base de datos se resume en la Figura 1. Las decisiones metodológicas responden a la necesidad de garantizar la coherencia del modelo, la estabilidad estadística y la validez en el uso de las transformaciones logarítmicos. La eliminación progresiva de *outliers* y casos inconsistentes permitió mitigar la presencia de valores extremos, comunes en bases censales. En ese sentido, la muestra conserva representatividad nacional e incluye negocios de diversas regiones y sectores económicos, garantizando una adecuada heterogeneidad del análisis.

**Figura 1. Procedimiento de selección, variables y análisis de datos**



Fuente: INDEC (2024).

### B. Definición y operacionalización de variables

La variable dependiente del modelo fue la condición empresarial (0 = Formal, 1 = Informal). De acuerdo con los criterios del INEI, una empresa formal es aquella con RUC activo y cumplimiento básico de obligaciones legales, mientras que se considera informal a aquella sin RUC o licencia municipal. Modelar la probabilidad de formalidad permite identificar los factores que mejor explican esta condición en las MYPEs, aportando diseño de políticas públicas focalizadas.

Las variables independientes se seleccionaron con base en estudios previos sobre informalidad empresarial (La Porta & Shleifer, 2014; Tokman, 2001) y en la información

disponible en el censo. Entre los factores estructurales considerados se encuentran: la ubicación geográfica de la empresa (Lima Metropolitana vs. provincias), la organización jurídica (persona natural vs. jurídica), el sector económico (producción, comercio o servicios), la tenencia del local (propio, alquilado o sin local fijo) y el tamaño de la empresa (número de trabajadores). Como factores estratégicos se incluyeron los gastos operativos anuales en remuneraciones, el valor agregado de producción, el volumen de ventas y el resultado de ejercicios. Todas estas variables fueron transformadas en escala logarítmica para normalizar su distribución y facilitar la interpretación estadística (Hair et al., 2010).

Las variables numéricas continuas fueron transformadas mediante logaritmos naturales para reducir asimetrías y, posteriormente, fueron centradas y escaladas utilizando la desviación estándar, con el fin de asegurar la estabilidad del modelo. Además, se comprobó la ausencia de colinealidad extrema entre predictores mediante el cálculo del factor de inflación de la varianza (VIF). Todos los valores fueron menores a 10, lo que indica un nivel aceptable de colinealidad y una independencia relativa entre los predictores<sup>4</sup>.

Cabe precisar que, para efectos del presente artículo, la formalidad fue definida operativamente como la tenencia de RUC activo y licencia municipal vigente, según los criterios del INEI. Sin embargo, existen casos mixtos — empresas con RUC, pero sin licencia, o viceversa— que fueron recodificados como informales durante el proceso de depuración de datos, a fin de mantener un criterio uniforme. Esta decisión metodológica responde a la necesidad de aplicar un criterio restrictivo que refleje el cumplimiento mínimo del marco legal.

### **C. Técnica de análisis estadístico**

El análisis estadístico comenzó con una descripción de la muestra, mediante frecuencias y medidas de tendencia central, lo que permitió caracterizar la distribución de empresas formales e informales. Posteriormente, se aplicó regresión logística binaria para estimar la probabilidad de formalidad ( $Y=0$ ) e identificar los factores que la explican. Esta técnica es ampliamente utilizada en ciencias sociales y económicas cuando la variable dependiente es categórica (Hosmer et al., 2013); permite calcular razones de momios (*odds ratios*) que muestran el efecto individual de cada predictor.

La elección de esta técnica responde a tres razones principales. La primera, porque permite interpretar de forma directa y transparente el efecto individual de cada predictor sobre la probabilidad de formalidad, a través del cálculo de momios (*odds ratios*), lo que facilita traducir los hallazgos en implicancias de política pública. La segunda, se debe a que guarda coherencia con el objetivo de esta investigación: identificar los factores que inciden significativamente en la condición de formalidad. La tercera, porque el modelo es estadísticamente robusto y metodológicamente trazable, ya que controla el efecto múltiple de los predictores y garantiza la replicabilidad en distintos contextos o períodos.

---

<sup>4</sup>Los valores de VIF obtenidos fueron: personal ocupado (1,2), resultado de ejercicio (2,7), remuneración (5,7), ventas (6,4) y valor agregado (9,4).

El modelo fue estimado mediante el método de máxima verosimilitud, incluyendo todas las variables en un solo paso, con el fin de controlar el efecto simultáneo de cada predictor. La significancia estadística se evaluó mediante el estadístico de Wald con un nivel de confianza del 95% ( $\alpha = 0,05$ ). Cabe precisar que la variable “organización jurídica” fue retirada del modelo final debido a valores extremos en los coeficientes ( $\beta \approx 19,682$ ), indicando un sobreajuste y separación casi perfecta entre las categorías, comprometiendo la estabilidad y la interpretabilidad del modelo (ver Anexo).

La calidad del modelo se evaluó con diversos indicadores. Primero, la prueba de ómnibus de coeficientes mostró que el modelo completo difiere significativamente de uno nulo ( $X^2 = 43829,236$ ;  $gl=11$ ;  $p<0,001$ ), confirmando que al menos uno de los coeficientes es estadísticamente significativo. Segundo, las medidas de pseudo  $R^2$  de Cox & Snell (0,221) y Nagelkerke (0,32) indicaron que el modelo explica entre el 22% y el 32% de la varianza en la formalidad empresarial. Además, la prueba de Hosmer-Lemeshow indicó un ajuste adecuado del modelo ( $X^2= 1269,523$ ;  $gl=8$ ;  $p<0,001$ ). Finalmente, la curva ROC arrojó un área bajo la curva (AUC) de 0,66, lo que representa una capacidad moderada para diferenciar entre empresas formales e informales (Bewick et al., 2005).

Por último, se analizó la precisión global mediante la matriz de clasificación, considerando un umbral de probabilidad estándar de 0,5 para asignar los casos como formales o informales. Dada la distribución desigual entre clases, también se exploró el ajuste del umbral para mejorar la sensibilidad hacia los casos informales. Todos los análisis se realizaron en el software estadístico SPSS versión 26.

### III. Resultados

El análisis descriptivo de la muestra evidenció que el 72,7% de las MYPEs registradas en el censo eran formales, mientras que el 27,3% restante correspondía a unidades informales. Esta distribución se alinea con la naturaleza de los censos económicos, que tienden a captar principalmente entidades con algún nivel de visibilidad institucional. Sin embargo, también evidencia la persistente presencia de informalidad en el tejido empresarial peruano (La Porta & Shleifer, 2014).

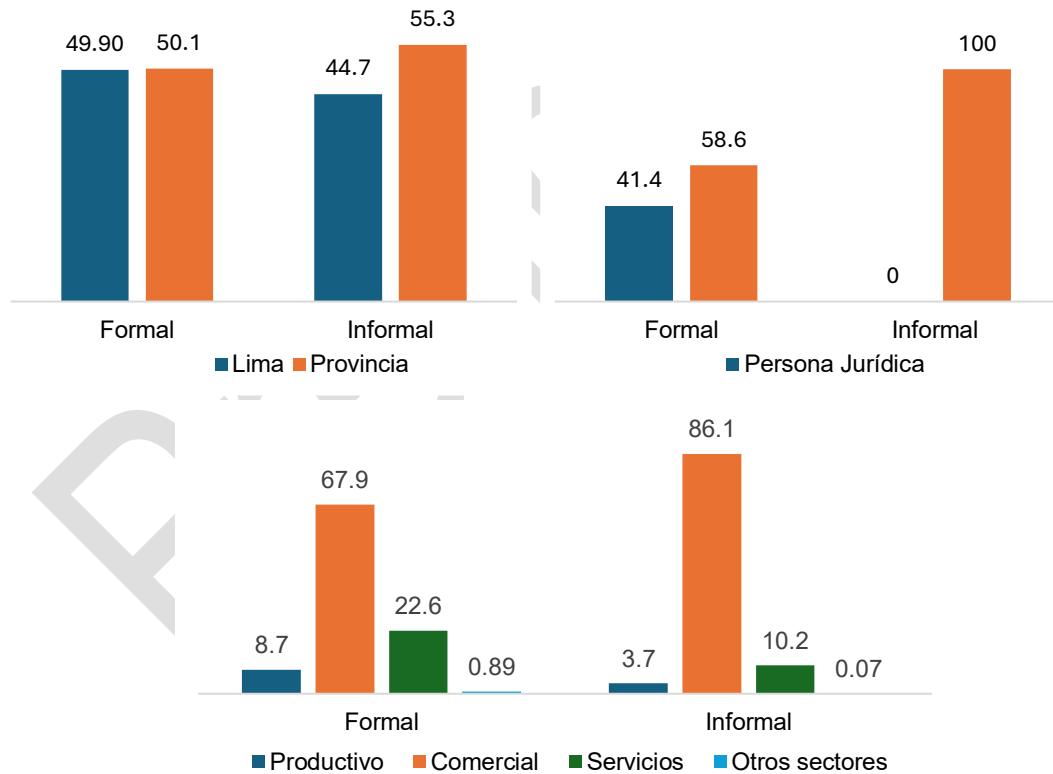
En términos sectoriales, el comercio representó la mayor proporción de las MYPEs (72,8%), seguido por el sector servicios (19,2%), actividades productivas (7,3%) y otros sectores (0,7%). Esta composición concuerda con las características de la economía peruana, donde el comercio minorista y los servicios predominan como formas de autoempleo y emprendimiento (Tokman, 2001). En cuanto a la ubicación geográfica, el 48,5% de las empresas operaban en Lima Metropolitana, mientras que el 51,5% restante se distribuía en otras regiones del país. Aunque esta distribución muestra una ligera sobrerepresentación de Lima, refleja la concentración de la actividad económica en la capital.

Respecto a la infraestructura operativa, el 52,4% de las empresas funcionaban en locales alquilados, el 45,9% en locales propios y el 1,7% carecía de un local. Esta última categoría incluye actividades informales como ventas ambulatorias o negocios domiciliarios no

diferenciados. En cuanto a la organización jurídica, el 69,9% de las empresas operaban como persona natural, mientras que solo el 30,1% estaban constituidas como persona jurídica, lo que sugiere que la formalización legal está estrechamente relacionada con la adopción de una estructura jurídica formal.

En la Figura 2, se observa que las empresas informales se concentran ligeramente más en provincias (55,3%) que en Lima (44,7%), mientras que las formales presentan una distribución más equilibrada (49,9% en Lima vs. 50,1% en provincias). En cuanto a la organización jurídica, todas las MYPEs informales son personas naturales, mientras que, entre las formales, el 58,6% son personas naturales y el 41,4% son personas jurídicas. Por sector económico, tanto las empresas formales como las informales se concentran principalmente en el comercio (67,9% y 86,1%, respectivamente), aunque las formales tienen una mayor participación en los servicios (22,6% frente a 10,2%) y en actividades productivas (8,7% frente a 3,7%).

**Figura 2.** Distribución de empresas formales e informales según ubicación geográfica, tipo de organización jurídica y sector económico

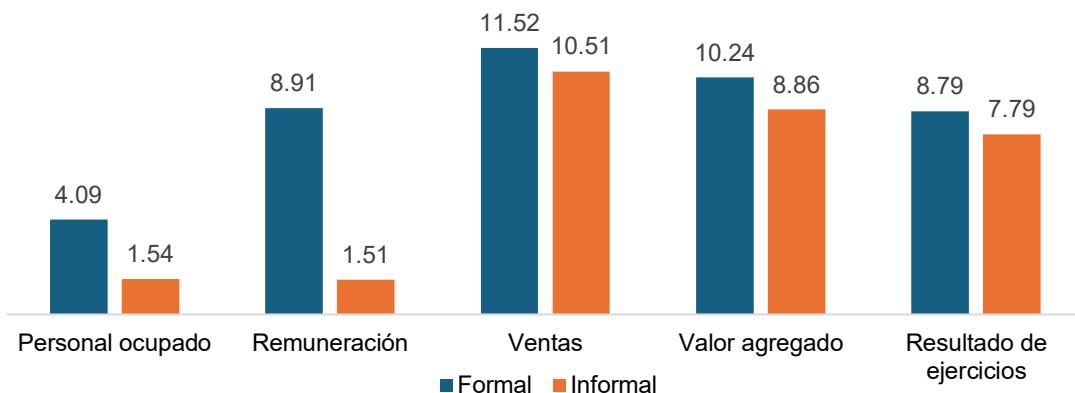


Fuente: INDEC (2024).

En cuanto a las variables continuas (Figura 3), las empresas formales presentan mayores promedios en todos los indicadores: número de personal ocupado (4,09 vs. 1,54), remuneración (8,91 vs. 1,51), ventas (11,52 vs. 10,51), valor agregado (10,24 vs. 8,86) y resultado de ejercicios (8,79 vs. 7,79). Estos valores reflejan que las empresas formales no

solo cuentan con mayor tamaño y capacidad productiva, sino que también obtienen mejores resultados financieros en comparación con sus contrapartes informales.

**Figura 3.** Distribución de empresas formales e informales según indicadores estratégicos



Fuente: INDEC (2024).

Para evaluar los factores determinantes de la formalidad, se estimó un modelo de regresión logística binaria, tomando como variable dependiente la condición empresarial (1=Informal, 0=Formal). El modelo resultó estadísticamente significativo ( $\chi^2 = 1269,523$ ;  $p < 0,001$ ), lo que confirma su capacidad explicativa superior frente a un modelo nulo. Se presentan los intervalos de confianza al 95% para cada *odds ratio* (OR), lo que permite evaluar la precisión de las estimaciones. Los coeficientes fueron expresados en términos de OR para facilitar su interpretación (véase la Tabla 1).

Los OR se interpretan en relación con las categorías de referencia: Lima (ubicación geográfica), sector productivo (actividad económica) y local fijo (tenencia de local). En el caso de las variables transformadas en logaritmos naturales (remuneración, ventas, valor agregado del producto, y resultado del ejercicio), los OR reflejan el efecto proporcional de un incremento relativo en dichas variables sobre la probabilidad de formalidad. Por ejemplo, un OR de 1,335 en ventas implica que un aumento del 1% en las ventas incrementa en 0,335 la probabilidad de formalidad, manteniendo constantes las demás variables.

**Tabla 1.** Modelo predictivo de la formalidad

Variable predictoría	Odds Ratio (Exp(B))	p-valor	I.C. 95%
Ubicación geográfica (Lima vs. provincias)	0,780	0,000	[0,760 – 0,800]
<i>Tipo de actividad económica</i>			
Servicios vs. Productivo	6,143	0,000	[4,160 – 9,072]

Variable predictora	Odds Ratio (Exp(B))	p-valor	I.C. 95%
Comercio vs. Productivo	9,878	0,000	[6,717 – 14,525]
Otro sector vs Productivo	6,073	0,000	[4,122 – 8,947]
Tamaño de la empresa (Número de empleados)	0,480	0,000	[0,472 – 0,488]
<i>Tenencia del local</i>			
Local alquilado vs Local fijo	0,601	0,000	[0,553 – 0,654]
Sin local fijo / otro vs Local fijo	0,404	0,000	[0,371 - 0,439]
Remuneración (Ln_remuneración)	0,888	0,000	[0,876 – 0,900]
Valor Agregado de Producción (Ln_valor_agregado)	0,508	0,000	[0,497 – 0,520]
Volumen de ventas (Ln_ventas)	1,335	0,000	[1,303– 1,368]
Resultado de ejercicio (Ln_resultado_ejercicio)	1,233	0,000	[1,219 – 1,247]

*Nota:* un odds ratio (OR) mayor a 1 indica que la variable aumenta la probabilidad del evento (en este caso, la formalidad), mientras que un OR menor a 1 la disminuye.

*Fuente:* elaboración propia.

Entre los principales resultados, destaca que el sector económico constituye uno de los predictores más influyentes: las MYPEs del sector comercio ( $OR \approx 9,878$ ), servicios ( $OR \approx 6,143$ ) y otros sectores ( $OR \approx 6,073$ ) presentan una probabilidad significativamente mayor de ser formales en comparación al sector productivo. En otras palabras, las MYPEs comerciales son casi diez veces más propensas a formalizarse que las productivas, mientras que las de servicios y otros sectores tienen una probabilidad aproximadamente de seis veces mayor.

Asimismo, el tamaño de la empresa, medido por el número de empleados, también se asoció significativamente con la formalización ( $OR \approx 0,480$ ). Por cada trabajador adicional se reduce la probabilidad de informalidad en un 52%, lo que confirma que las empresas con mayor cantidad de trabajadores tienden a regularizarse. Este hallazgo coincide con estudios que vinculan el crecimiento organizacional con mayores incentivos y necesidades de cumplimiento normativo (Günther & Launov, 2012).

En cuanto a la infraestructura operativa, las empresas que operaban en locales alquilados presentaron menor probabilidad de ser informales que aquellas con local propio ( $OR \approx 0,601$ ), y las que carecían de local formal mostraron aún menor probabilidad de formalidad ( $OR \approx 0,404$ ). Aunque esta relación puede parecer contraintuitiva, podría explicarse por el hecho de que los espacios alquilados en entornos formales (como mercados o galerías) conllevan mayores exigencias legales y una mayor visibilidad ante las autoridades fiscales. Por

el contrario, muchas empresas con locales propios funcionan en entornos domésticos o semiinformales donde la fiscalización es limitada.

Respecto al desempeño económico, los mayores niveles de ventas ( $OR \approx 1,335$ ) y el resultado del ejercicio ( $OR \approx 1,233$ ) se asociaron con una mayor probabilidad de formalización. Esto sugiere que las MYPEs con mejor rendimiento financiero buscan regularizarse, posiblemente para acceder a beneficios como créditos o contratos formales. Este resultado es coherente con la literatura que plantea que la formalidad puede ser una condición necesaria para integrarse a cadenas de valor o escalar productivamente (World Bank, 2020). Sin embargo, el valor agregado ( $OR \approx 0,508$ ) y las remuneraciones ( $OR \approx 0,888$ ) mostraron asociaciones inversas, lo que sugiere que un mayor gasto operativo no necesariamente se traduce en mayor formalidad.

Estos resultados indican que, si bien el crecimiento de ingresos puede impulsar la formalización, no todos los indicadores productivos siguen la misma lógica, lo que apunta a posibles estrategias deliberadas para operar fuera del marco regulatorio (de Soto, 1986). La ubicación geográfica también mostró un efecto significativo, aunque moderado ( $OR \approx 0,78$ ), indicando que operar en Lima reduce en un 22% la probabilidad de informalidad en comparación con provincias. Este efecto, aunque relevante, es menor respecto al de las variables económicas y estructurales.

En cuanto al ajuste del modelo, la prueba de Hosmer-Lemeshow ( $\chi^2 \approx 1269,52$ ;  $p < 0,001$ ) mostró discrepancias entre los valores observados y los predichos. Sin embargo, debido al gran tamaño de la muestra, incluso diferencias pequeñas pueden resultar estadísticamente significativas, por lo que este resultado no invalida el modelo, aunque sí sugiere oportunidades de mejora mediante la incorporación de variables adicionales o interacciones.

La curva ROC arrojó un área bajo la curva (AUC) de  $0,660^5$ , lo que representa una capacidad discriminativa moderada (Hosmer et al., 2013). Es decir, el modelo logra distinguir adecuadamente entre empresas formales e informales en aproximadamente el 66% de los casos, un desempeño satisfactorio para un modelo con múltiples predictores en contextos empresariales. Por su parte, el pseudo  $R^2$  de Nagelkerke fue de 0,320, lo que indica que el modelo explica alrededor del 32% de la varianza en la condición de formalidad, cifra comparable a la encontrada en estudios similares donde intervienen múltiples factores contextuales difíciles de capturar empíricamente (Günther & Launov, 2012).

En términos de precisión, el modelo clasificó correctamente el 78,6% de los casos, con una sensibilidad del 91,9% (identificación correcta de empresas formales) y una especificidad del 43,3% (detección de informales). Esta asimetría responde al desbalance de clases en la muestra. El análisis principal se realizó con un punto de corte estándar de 0,5 para predecir la formalidad; adicionalmente, se probó un ajuste alternativo (corte 0,3) para evaluar la utilidad de detección en empresas informales, cuyos resultados se presentan en el Anexo.

---

<sup>5</sup> El área bajo la curva (AUC) mide la capacidad del modelo para distinguir entre empresas formales e informales. Un AUC de 0,5 indica que el modelo no discrimina mejor que el azar, mientras que valores cercanos a 1 reflejan una excelente capacidad de clasificación.

#### IV. Discusión

Los resultados confirman que la formalización empresarial está fuertemente determinada por factores internos como el tamaño, la infraestructura operativa, el desempeño económico y, en menor medida, la ubicación. También ponen de relieve que la existencia de ingresos elevados no garantiza por sí sola la formalización y que los factores contextuales, como el sector y la región, aunque estadísticamente significativos, tienen menor peso en comparación con los factores estructurales y estratégicos. En conjunto, estos hallazgos ofrecen una visión integral sobre los determinantes de la formalidad en las MYPEs peruanas.

Los hallazgos reafirman la importancia de los factores estructurales y estratégicos en la probabilidad de formalización de las MYPEs peruanas. En consonancia con lo propuesto en el marco teórico, se identificó que el tamaño de la empresa, la tenencia del local, el valor agregado de la producción, el resultado del ejercicio y el volumen de ventas son los predictores más significativos de la formalidad. Estos resultados no solo son consistentes con la literatura previa, sino que también amplían el conocimiento empírico al aplicar un modelo predictivo sobre una base censal nacional, lo que otorga mayor robustez a las inferencias realizadas.

El tamaño de la empresa, medido por el número de empleados, mostró una relación inversa con la informalidad ( $OR \approx 0,480$ ), lo que indica que cada trabajador adicional reduce significativamente la probabilidad de operar fuera del marco legal. Este hallazgo reafirma que el crecimiento organizacional actúa como catalizador de la formalización, como lo documentan estudios en África y América Latina (Aberra et al., 2022; World Bank, 2020). Las empresas con mayor cantidad de personal tienden a regularizarse para poder contratar formalmente, acceder a servicios financieros e integrarse a cadenas de valor.

Asimismo, operar desde un local alquilado, más que desde un local propio, se asoció con mayor formalidad ( $OR \approx 0,601$ ). Este resultado puede explicarse por la presión contractual, la mayor visibilidad ante las autoridades fiscales y la inserción en espacios comerciales regulados. En el ámbito del desempeño, el valor agregado y el resultado del ejercicio fueron predictores robustos de formalización ( $OR = 0,508$  y  $OR = 1,233$ , respectivamente), lo que sugiere que las empresas más productivas perciben mayores beneficios o menores costos relativos al cumplimiento normativo. Este resultado coincide con el enfoque del World Bank (2020), que destaca que la participación en mercados competitivos y regulados impulsa a las empresas a formalizarse como condición para crecer.

De manera consistente, el volumen de ventas mostró una asociación positiva con la formalidad ( $OR = 1,335$ ), indicando que las empresas con mayores ingresos tienden a regularizarse, posiblemente para acceder a mercados más amplios, financiamiento y contratos formales. Este hallazgo refuerza la idea de que el crecimiento económico interno puede convertirse en un incentivo para el cumplimiento normativo, aunque la decisión de formalizarse puede verse modulada por factores como los costos regulatorios, percepción de corrupción o la limitada fiscalización (de Soto, 1986; Robles & Ambriz, 2023).

La ubicación geográfica (Lima vs. provincia) y el sector económico (comercio y servicios frente a productivo) también resultaron estadísticamente significativos, aunque su efecto fue menor en comparación con los factores estructurales. Esto sugiere que las diferencias regionales y sectoriales reflejan principalmente la composición empresarial más que un efecto causal directo. Estos resultados coinciden con Buitrago et al. (2024), quienes sostienen que la calidad institucional y los marcos regulatorios tienen mayor peso que las características sectoriales o regionales en la formalización.

En cuanto al desempeño metodológico, el modelo predictivo alcanzó una capacidad discriminativa aceptable ( $AUC = 0,660$  y una precisión global del 78,6%, comparable a la de estudios similares con muestras grandes y múltiples predictores (Hosmer et al., 2013). Sin embargo, se identificó una asimetría en la capacidad del modelo para predecir informalidad frente a formalidad, con menor sensibilidad hacia las empresas informales. Esto evidencia la necesidad de ajustar umbrales o incorporar nuevas variables que puedan capturar con mayor precisión la heterogeneidad del segmento informal, como el acceso a financiamiento, el uso de tecnología o factores de comportamiento del emprendedor.

#### **A. Limitaciones y futuras líneas de investigación**

La presente investigación tiene algunas limitaciones que deben considerarse al interpretar los resultados. En primer lugar, aunque la regresión logística binaria fue seleccionada por su capacidad de interpretar de manera directa y transparente los efectos de cada predictor, su coherencia con el objetivo del estudio y su robustez metodológica, no se realizó una comparación con modelos alternativos como árboles de decisión, *random forest* o máquinas de soporte vectorial. Explorar estas técnicas en futuras investigaciones podría permitir evaluar posibles mejoras en la capacidad predictiva sin sacrificar la interpretabilidad.

En segundo lugar, si bien la variable ubicación geográfica fue incluida como predictor, no se efectuó un análisis de sensibilidad que verificara la robustez de los coeficientes estimados en submuestras específicas. Por ejemplo, no se desarrollaron modelos separados para empresas ubicadas en Lima Metropolitana y provincias. Este análisis habría permitido identificar posibles variaciones estructurales en los determinantes de la formalidad. Se recomienda que futuras investigaciones incorporen análisis multigrupo o segmentaciones por ubicación geográfica a fin de evaluar la consistencia de los efectos en diferentes contextos.

En tercer lugar, el modelo no incluye variables institucionales o del entorno, tales como la carga tributaria local, las facilidades para la formalización a nivel municipal o los niveles de corrupción percibida. La exclusión de estos factores responde a la falta de datos con desagregación subnacional en el V Censo Nacional Económico 2022. Aunque el enfoque estructural y estratégico adoptado permite capturar dimensiones internas de la empresa, estudios previos han demostrado que los indicadores de burocracia, calidad institucional y regulaciones locales se asocian fuertemente con la informalidad. Por lo tanto, incorporar estas variables en futuros modelos permitiría enriquecer el análisis y comprender con mayor profundidad las dinámicas contextuales que condicionan la formalización empresarial.

Finalmente, al trabajar con una base censal de corte transversal, no fue posible analizar dinámicas temporales ni establecer relaciones causales entre los predictores y la formalidad

empresarial. El uso de encuestas longitudinales en investigaciones futuras podría permitir capturar trayectorias de formalización y evaluar los efectos diferidos de las políticas públicas. A pesar de estas limitaciones, el estudio aporta evidencia sólida sobre los factores estructurales y estratégicos que inciden en la formalidad empresarial, ofreciendo una herramienta útil para el diseño de políticas públicas orientadas a promover la regularización de las MYPEs en el Perú.

### Conclusiones

Los resultados de este artículo evidencian que la formalidad en las MYPEs peruanas está principalmente determinada por factores internos, como el tamaño de la empresa, la tenencia del local, el valor agregado de la producción, las ventas, y los resultados de ejercicio. Estos elementos explican de manera significativa la probabilidad de formalización, superando el peso relativo de variables contextuales como ubicación geográfica o sector económico.

Desde el punto de vista práctico, estos hallazgos ofrecen insumos valiosos para el diseño de políticas públicas. La identificación de perfiles empresariales con alta probabilidad de informalidad permite focalizar los esfuerzos estatales en segmentos específicos, particularmente en microempresas con figura de persona natural, baja dotación de personal, reducida producción o sin local comercial definido. Por ejemplo, entidades como SUNAT y PRODUCE podrían utilizar modelos similares para diseñar campañas de formalización dirigidas, establecer incentivos fiscales progresivos o desarrollar programas de acompañamiento empresarial. Asimismo, los gobiernos locales pueden facilitar el acceso a espacios con condiciones asequibles para negocios en proceso de transición hacia la formalidad.

En el ámbito académico, se recomienda complementar el modelo con variables derivadas de encuestas longitudinales o paneles que permitan capturar trayectorias empresariales y efectos de largo plazo. Explorar modelos de aprendizaje automático capaces de manejar grandes volúmenes de datos y relacionales no lineales podría mejorar la capacidad predictiva y la detección de patrones complejos. Además, el análisis geoespacial de la informalidad, integrando datos de ubicación, infraestructura urbana y conectividad digital, y el estudio de la interacción entre género del emprendedor y la decisión de formalizarse, también ofrecería aportes valiosos a la literatura actual.

Finalmente, este estudio confirma que la formalización de las MYPEs peruanas no depende únicamente del entorno regulatorio, sino que está fuertemente condicionada por decisiones empresariales vinculadas a su estructura y desempeño. El modelo propuesto constituye un aporte metodológico y empírico relevante, ofreciendo una herramienta útil para la formulación de políticas más eficaces y focalizadas en la reducción de la informalidad empresarial.

### Declaración de ética

Este artículo de investigación no realizó trabajo con una persona o grupos de personas para la generación de datos empleados en la metodología, por tanto, no requirió ni obtuvo un aval de Comité de Ética para su realización.

## Referencias

- Aberra, A., Aga, G., Jolevski, F., & Karalashvili, N. (2022). *Understanding Informality: Comprehensive Business-Level Data and Descriptive Findings* [Policy Research Working Paper, 10208]. World Bank Group. <https://www.enterprisesurveys.org/content/dam/enterprisesurveys/documents/research-1/Understanding%20informality.pdf>
- Alvarado Lagunas, E. (2021). Factores condicionantes en la creación informal de nanoempresas: Evidencia experimental en Monterrey, México. *Contaduría y Administración*, 66(3). <https://doi.org/10.22201/fca.24488410e.2021.2848>
- Bewick, V., Cheek, L., & Ball, J. (2005). Statistics Review 14: Logistic regression. *Critical Care*, 9(1), 112–118. <https://doi.org/10.1186/cc3045>
- Bravo-Ortega, C., Bustamante, C., Egana, P., Symmes, F., & Sexton, J. (2024). The Formalization of Microenterprises in Middle-Income Countries: Informal Institutions as a Mechanism to Address Institutional Incongruence. *Management Research*, 22(2). <https://doi.org/10.1108/mrijam-04-2023-1417>
- Buitrago, R., García-Suaza, A., & Garzón, J. (2024). Exploratory Analysis of the Determinants of Informality in Emerging and Frontier Economies: An Institutional Approach via PLS-SEM. *Development Studies Research*, 11(1). <https://doi.org/10.1080/21665095.2023.2291004>
- de Soto, H. (1986). *El otro sendero: La revolución informal*. Instituto Libertad y Democracia.
- Günther, I., & Launov, A. (2012). Informal Employment in Developing Countries: Opportunity or Last Resort? *Journal of Development Economics*, 97(1), 88–98. <https://doi.org/10.1016/j.jdeveco.2011.01.001>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Pearson.
- Hernández, R., Fernández, C., & Baptista, P. (2014). *Metodología de la investigación* (6.<sup>a</sup> ed.). McGraw-Hill.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (3rd ed.). Wiley.
- Instituto Nacional de Estadística e Informática (INEI). (2024). *V Censo Nacional Económico (CENEC 2022): Resultados definitivos*. <https://www.gob.pe/institucion/inei/informes-publicaciones/5638115-v-censo-nacional-economico-v-cenec-resultados-definitivos>
- Kede, H., Houehanou, A., & Nguimkeu, P. (2023). The Impact of Innovation on Formalization and Job Creation of Informal SMEs in Cameroon. *Econometric Society*. [https://www.econometricsociety.org/event\\_papers/download/265/186/1/kedeet.al.2023innovation.pdf](https://www.econometricsociety.org/event_papers/download/265/186/1/kedeet.al.2023innovation.pdf)
- La Porta, R., & Shleifer, A. (2014). Informality and Development. *Journal of Economic Perspectives*, 28(3), 109–126. <https://doi.org/10.1257/jep.28.3.109>

- Loayza, N. V. (1997). The Economics of the Informal Sector: A Simple Model and Some Empirical Evidence from Latin America. *Carnegie-Rochester Conference Series on Public Policy*, 45, 129–162. [https://doi.org/10.1016/S0167-2231\(96\)00021-8](https://doi.org/10.1016/S0167-2231(96)00021-8)
- Mehjabeen, M., & Noor, S. (2023). Motivating Factors Affecting Formalization of Small Informal Enterprises: Evidence from an Emerging Economy. *Business Strategy & Development*, 7(3). <https://doi.org/10.1002/bsd2.426>
- Robles, D., & Ambriz, R. (2024). Informal Employment, the Tertiary Sector, and the Gross Domestic Product: A Structural Equations Model for the Mexican Economy. *Economies*, 12(11), 301. <https://doi.org/10.3390/economies12110301>
- Rodríguez Medina, O., Alvarado Lagunas, E., & Sánchez Silva, M. (2023). Informalidad y nanoempresas en localidades perimetropolitanas de la Ciudad de México. *Problemas del Desarrollo*, 54(212), 55–77. <https://doi.org/10.22201/iiec.20078951e.2023.212.69898>
- Superintendencia Nacional de Administración Tributaria (SUNAT). (2025). *Características de la micro y pequeñas empresas*. <https://www.sunat.gob.pe/orientacion/mypes/caracteristicas-microPequenaEmpresa.html>
- Tokman, V. E. (2001). Integrating the Informal Sector into the Modernization Process. *SAINS Review*, 21(1), 45–60. <https://doi.org/10.1353/sais.2001.0027>
- Top, M., & Zhi, Y. (2023). Growing an Informal Digital Micro-Enterprise: The Role of Individual Entrepreneurial Orientation and Social Capital. *International Journal of Innovation Management*, 27(9–10). <https://doi.org/10.1142/S1363919623500536>
- World Bank. (2020). World Development Report 2020: Trading for Development in the Age of Global Value Chains. World Bank Publications. <https://www.worldbank.org/en/publication/wdr2020>

**Anexo****Tabla A1.** Comparación de los modelos de regresión logística binaria para predecir la formalidad

Variable predictora	Modelo 1 (con organización jurídica)			Modelo 2 (sin organización jurídica)		
	Odds Ratio (Exp(B))	p-valor	I.C. 95%	Odds Ratio (Exp(B))	p-valor	I.C. 95%
Organización jurídica (Persona Natural vs Persona jurídica)	352950543,6	,892				
Ubicación geográfica (Lima vs. provincias)	0,862	,000	[0,839 – 0,886]	0,780	0,000	[0,760 – 0,800]
<i>Tipo de actividad económica</i>						
Servicios vs. Productivo	1,797	,000	[4,160 – 9,072]	6,143	0,000	[4,160 – 9,072]
Comercio vs. Productivo	1,720	,017	[6,717 – 14,525]	9,878	0,000	[6,717 – 14,525]
Otro sector vs Productivo	1,332	,025	[4,122 – 8,947]	6,073	0,000	[4,122 – 8,947]
Tamaño de la empresa (Número de empleados)	0,477	,000	[0,472 – 0,488]	0,480	0,000	[0,472 – 0,488]
<i>Tenencia del local</i>						
Local alquilado vs Local fijo	0,706	0,000	[0,553 – 0,654]	0,601	0,000	[0,553 – 0,654]
Sin local fijo / otro vs Local fijo	0,489	0,000	[0,371 - 0,439]	0,404	0,000	[0,371 - 0,439]
Remuneración (Ln_remuneración)	0,919	0,000	[0,876 – 0,900]	0,888	0,000	[0,876 – 0,900]
Valor Agregado de Producción (Ln_valor_agregado)	0,605	0,000	[0,497 – 0,520]	0,508	0,000	[0,497 – 0,520]
Volumen de ventas (Ln_ventas)	1,617	0,000	[1,303– 1,368]	1,335	0,000	[1,303– 1,368]
Resultado de ejercicio (Ln_resultado_ejercicio)	1,185	0,000	[1,219 – 1,247]	1,233	0,000	[1,219 – 1,247]

Fuente: INDEC (2024).

**Tabla A2.** Comparación de indicadores de los modelos de regresión logística binaria para predecir la formalidad

Métrica	Modelo 1 (con organización jurídica)	Modelo 2 (sin organización jurídica)
Pseudo R <sup>2</sup> (Nagelkerke)	0,452	0,320
AUC	0,497	0,660
Clasificación global (corte 0,5)	80,1	78,6
Sensibilidad (corte 0,5)	90,1 (formal) y 53,4 (informal)	91,9 (formal) y 43,3 (informal)
Especificidad (corte 0,5)	53,4%	43,3%
Clasificación global (corte 0,3)	72,1	71,2
Sensibilidad (corte 0,3)	68,2 (formal) y 82,8 (informal)	70,6 (formal) y 72,9 (informal)
Especificidad (corte 0,3)	82,8%	72,9%

Fuente: INDEC (2024).