

Informe de Análisis Bivariado del Régimen de Afiliación

1. Introducción

El análisis bivariado constituye una etapa fundamental dentro del proceso exploratorio de datos, ya que permite identificar posibles relaciones entre pares de variables y comprender cómo interactúan entre sí dentro de un conjunto de información.

En este caso, se buscó examinar la relación entre diferentes características demográficas, socioeconómicas y de afiliación, como el género, la edad, la localización geográfica, el nivel del Sisbén y el régimen de afiliación.

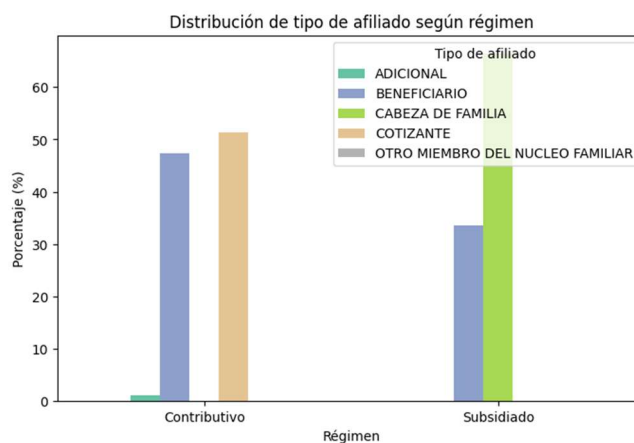
A través de la prueba de independencia Chi-cuadrado y la medida de intensidad Cramér's V, se evaluó tanto la significancia estadística como la fuerza de asociación entre las variables, con el fin de reconocer patrones estructurales y desigualdades potenciales en la distribución de la población afiliada.

Los resultados obtenidos ofrecen una base descriptiva sólida para la interpretación de los factores que inciden en el comportamiento del sistema de salud y en la caracterización social de sus beneficiarios.

2. Régimen vs Tipo de Afiliado Contributivo

- Distribución porcentual (%):

Régimen	ADICIONAL	BENEFICIARIO	CABEZA DE FAMILIA	COTIZANTE	OTRO MIEMBRO
Contributivo	1.14	47.45	0.00	51.41	0.00
Subsidiado	0.00	33.51	66.49	0.00	0.00



- Resultados estadísticos:

- $\chi^2 = 1,045,368.142$

- $gl = 4$
- $p\text{-valor} = 0.0000$
- Cramér's $V = 0.940 \rightarrow$ Asociación extremadamente fuerte y significativa.

- Interpretación:

La relación entre el régimen de afiliación y el tipo de afiliado es altamente significativa y presenta una fuerza de asociación casi perfecta (Cramér's $V = 0.94$).

El régimen contributivo concentra la totalidad de los cotizantes y adicionales, mientras que el régimen subsidiado agrupa casi exclusivamente a los beneficiarios y cabezas de familia.

Este patrón evidencia una dependencia estructural entre ambas variables, pues el tipo de afiliado está determinado por el régimen al que pertenece la persona.

En términos prácticos, el resultado refleja la lógica normativa del sistema de salud:

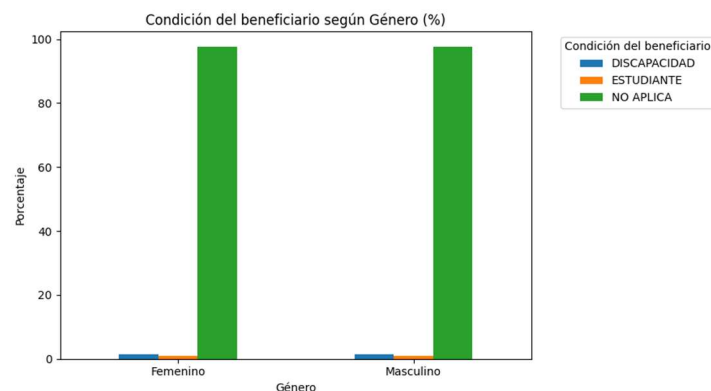
- Los cotizantes y adicionales pertenecen al régimen contributivo, financiado por aportes laborales.
- Los beneficiarios y cabezas de familia corresponden al régimen subsidiado, orientado a población sin capacidad de pago.

Por tanto, aunque la asociación estadística es extremadamente fuerte, su interpretación debe entenderse en un contexto determinístico y no probabilístico, derivado del diseño institucional del sistema de afiliación.

3. Género vs Condición del Beneficiario

- Distribución porcentual (%):

Género	DISCAPACIDAD	ESTUDIANTE	NO APLICA
Femenino	1.45	1.01	97.54
Masculino	1.51	0.87	97.62



- Resultados estadísticos:

- $\chi^2 = 96.731$
- $gl = 2$
- $p < 0.001$
- Cramér's $V = 0.007 \rightarrow$ Existe asociación significativa (muy débil)

- Interpretación:

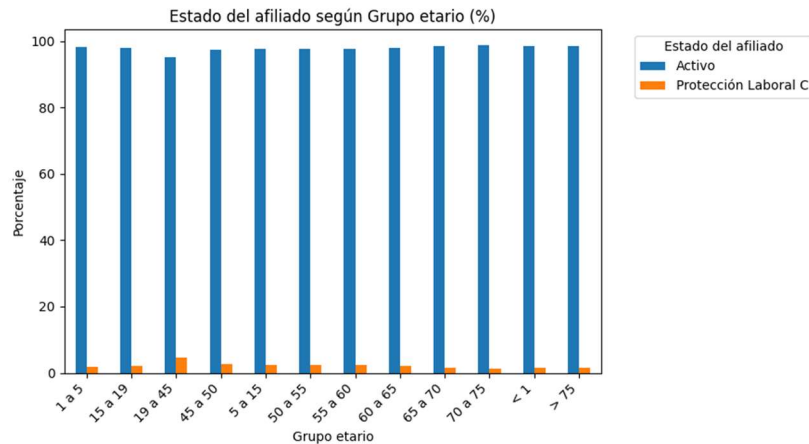
La prueba Chi-cuadrado confirma una asociación estadísticamente significativa, aunque el valor de Cramér's V (0.007) indica que la relación es muy débil.

Los porcentajes son prácticamente equivalentes entre géneros, lo que implica que, si bien existe una dependencia estadística, esta no tiene relevancia práctica en términos de comportamiento diferencial entre hombres y mujeres.

4. Grupo Etario vs Estado del Afiliado

- Distribución porcentual (%):

Grupo Etario	Activo	Protección Laboral C
< 1	98.48	1.52
1 a 5	98.18	1.82
5 a 15	97.51	2.49
15 a 19	98.02	1.98
19 a 45	95.24	4.76
45 a 50	97.31	2.69
50 a 55	97.61	2.39
55 a 60	97.67	2.33
60 a 65	97.86	2.14
65 a 70	98.38	1.62
70 a 75	98.63	1.37
> 75	98.56	1.44



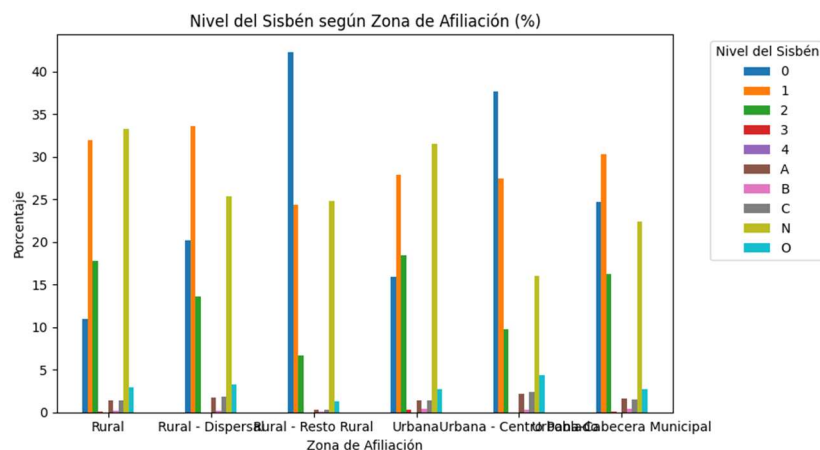
- Resultados estadísticos:
 - $\chi^2 = 8912.670$
 - $gl = 11$
 - $p < 0.001$
 - Cramér's $V = 0.071 \rightarrow$ Existe asociación significativa (débil a moderada).
- Interpretación:

La relación entre edad y estado del afiliado es estadísticamente significativa, mostrando que los grupos de edad productiva (19–45 años) presentan un mayor porcentaje de afiliados activos, mientras que la categoría Protección Laboral C aumenta ligeramente en grupos mayores.

El valor de Cramér's V (0.071) confirma una asociación débil a moderada, pero consistente con las dinámicas laborales del sistema de salud.

5. Zona de Afiliación vs Nivel del Sisbén

- Distribución porcentual (%):



- Resultados estadísticos:
 - $\chi^2 = 48216.202$
 - $gl = 45$
 - $p < 0.001$
 - Cramér's $V = 0.074 \rightarrow$ Existe asociación significativa (moderada).

- Interpretación:

La relación entre zona de afiliación y nivel del Sisbén muestra diferencias importantes entre áreas rurales y urbanas.

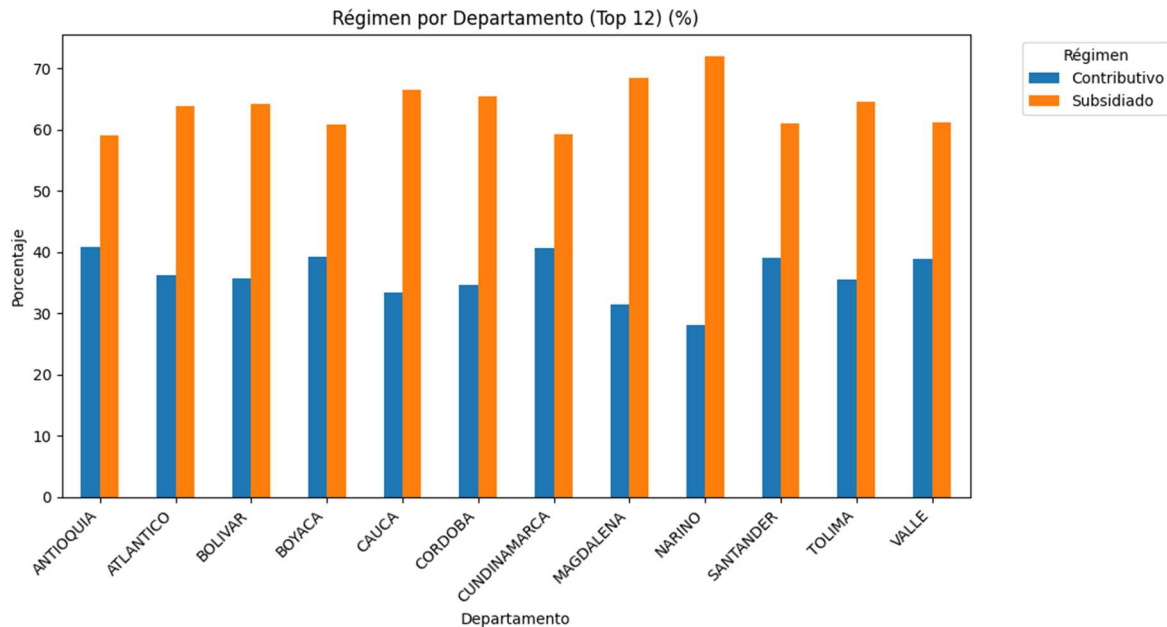
Las zonas rurales y dispersas concentran mayores porcentajes en los niveles 1 y N, reflejando condiciones socioeconómicas más bajas, mientras que las urbanas presentan una distribución más equilibrada entre niveles.

El valor de Cramér's V (0.074) indica una asociación significativa y de intensidad moderada, confirmando disparidades territoriales en la clasificación socioeconómica.

6. Departamento vs Régimen

- Distribución porcentual (%):

Departamento	Contributivo	Subsidiado
Antioquia	40.84	59.16
Atlántico	36.22	63.78
Bolívar	35.78	64.22
Boyacá	39.20	60.80
Cauca	33.48	66.52
Córdoba	34.57	65.43
Cundinamarca	40.66	59.34
Magdalena	31.56	68.44
Nariño	28.04	71.96
Santander	39.03	60.97
Tolima	35.52	64.48
Valle	38.83	61.17



- Resultados estadísticos:

- $\chi^2 = 6683.259$
- $gl = 11$
- $p < 0.001$
- Cramér's V = 0.075 → Existe asociación significativa.

- Interpretación:

Se evidencia una asociación significativa entre el departamento de residencia y el régimen de afiliación.

Departamentos como Antioquia, Cundinamarca y Valle del Cauca presentan una proporción más equilibrada entre regímenes, mientras que regiones del suroccidente y la costa caribe muestran una mayor concentración en el régimen subsidiado.

El valor de Cramér's V (0.075) confirma una asociación débil pero relevante, lo que indica que el tipo de régimen está influido parcialmente por las condiciones geográficas y socioeconómicas del territorio.

Conclusiones

Se identificaron asociaciones estadísticamente significativas en la mayoría de los cruces analizados, lo que demuestra que las variables demográficas, geográficas y socioeconómicas presentan interdependencias relevantes dentro del conjunto de datos.

El cruce entre Régimen y Tipo de Afiliado evidenció una asociación extremadamente fuerte (Cramér's V = 0.940), lo que refleja una dependencia estructural entre ambas variables.

Este resultado confirma que la afiliación al régimen contributivo o subsidiado determina directamente la categoría del afiliado (cotizante, beneficiario o cabeza de familia), coherente con la normatividad del sistema de salud colombiano.

Las relaciones entre Departamento y Régimen (Cramér's $V = 0.075$) y entre Zona de Afiliación y Nivel del Sisbén (Cramér's $V = 0.074$) mostraron asociaciones moderadas, que evidencian diferencias territoriales y socioeconómicas marcadas.

Los departamentos y zonas rurales presentan mayores proporciones de afiliados en regímenes subsidiados o en niveles bajos del Sisbén, reflejando desigualdades estructurales en el acceso y cobertura del sistema.

El cruce Grupo Etario \times Estado del Afiliado (Cramér's $V = 0.071$) indicó una asociación débil a moderada, pero estadísticamente significativa, consistente con la dinámica laboral y demográfica: las personas en edad productiva se concentran en la categoría de "activos", mientras que los mayores tienden a ubicarse en "Protección Laboral C".

La relación Género \times Condición del Beneficiario (Cramér's $V = 0.007$), aunque estadísticamente significativa, presentó una intensidad muy baja, lo que sugiere que las diferencias entre hombres y mujeres en esta variable son mínimas y no tienen un peso estructural dentro del sistema de afiliación.

En conjunto, los resultados permiten concluir que las principales fuentes de variabilidad en la base de datos provienen de factores institucionales (régimen), territoriales (departamento, zona de afiliación) y socioeconómicos (nivel del Sisbén), más que de diferencias de género o edad.

Esto revela un sistema de afiliación fuertemente condicionado por el contexto geográfico y las condiciones socioeconómicas, lo cual tiene implicaciones relevantes para la equidad en salud y la formulación de políticas públicas.

Finalmente, el uso complementario del test Chi-cuadrado y del coeficiente de Cramér's V permitió no solo determinar la existencia de relaciones significativas, sino también evaluar su magnitud y relevancia práctica.

Informe de Análisis Multivariado del Régimen de Afiliación

1. Introducción

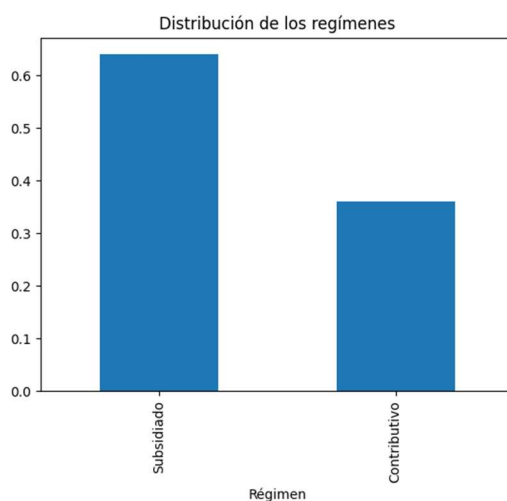
El presente análisis busca identificar los factores que explican la pertenencia de los individuos a los distintos regímenes del sistema de salud (Subsidiado y Contributivo), utilizando un conjunto de variables categóricas relacionadas con el tipo de afiliación, características socioeconómicas y entidad administradora.

Dado que todas las variables son categóricas, se aplicaron métodos estadísticos y de aprendizaje interpretables como:

- **Coefficiente de asociación (Cramer's V)** para determinar la fuerza de relación entre variables.
- **Análisis multivariado (SVD sobre codificación One-Hot)** para explorar patrones de agrupamiento.
- **Árbol de decisión** para identificar las reglas explicativas del régimen.
- **Análisis descriptivo** de la distribución general de los regímenes.

2. Distribución general de los regímenes

La base de datos presenta una mayor proporción de afiliados al régimen subsidiado ($\approx 64\%$) frente al contributivo ($\approx 36\%$).



- El régimen subsidiado tiene una participación predominante, lo cual refleja una mayor cobertura de la población en condiciones de vulnerabilidad o con bajos niveles del Sisbén.

- La menor proporción contributiva indica una base reducida de trabajadores formales o personas con capacidad de cotizar.
- Este patrón es coherente con la composición socioeconómica de los grupos y con los resultados del árbol de decisión, donde las variables económicas y de afiliación explican fuertemente el régimen.

3. Matriz de asociación (Cramer's V)

El coeficiente de asociación mostró las siguientes relaciones más fuertes con el régimen:

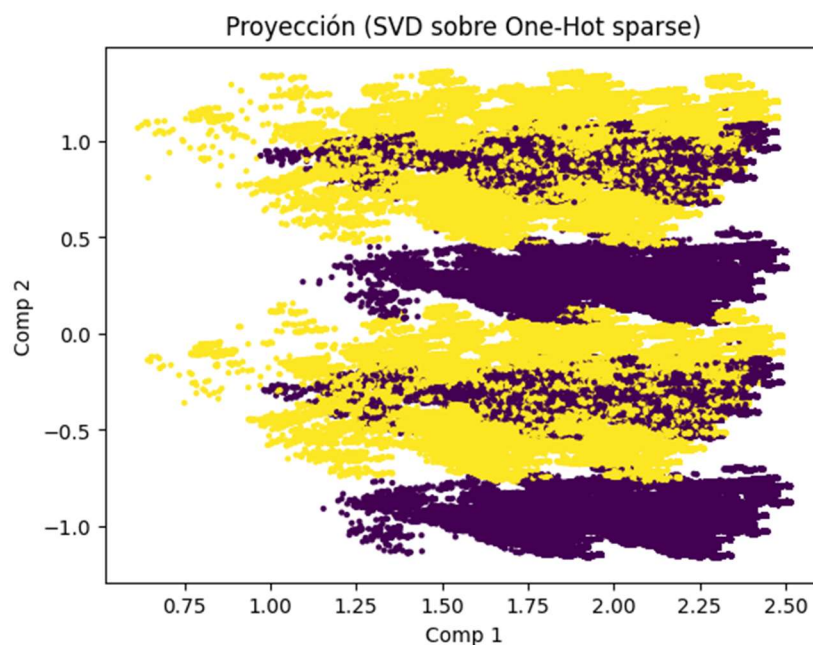
θ	
Código de la entidad	0.999983
Tipo de afiliado	0.766364
Nombre de la entidad	0.673055
Nivel del Sisbén	0.578552
Estado del afiliado	0.217868
Municipio	0.134788
Zona de Afiliación	0.125388
Condición del beneficiario	0.121343
Grupo etario	0.117238
Departamento	0.092710
Género	0.008943

Variable	Cramer's	Interpretación
Código de la entidad	0.9999	Asociación casi perfecta: las EPS operan prácticamente bajo un único régimen.
Tipo de afiliado	0.766	Fuerte relación: los cotizantes y miembros familiares pertenecen al contributivo, mientras los beneficiarios y cabezas de familia al subsidiado.
Nombre de la entidad	0.673	Alta relación institucional: refuerza la segmentación por EPS.
Nivel del Sisbén	0.579	Fuerte relación socioeconómica: niveles bajos → subsidiado.
Estado del afiliado / Zona / Condición del beneficiario	0.12–0.21	Relaciones secundarias.
Género / Departamento / Grupo etario	<0.10	Sin influencia relevante.

El régimen está determinado principalmente por variables institucionales (entidad) y socioeconómicas (tipo de afiliado, nivel del Sisbén).

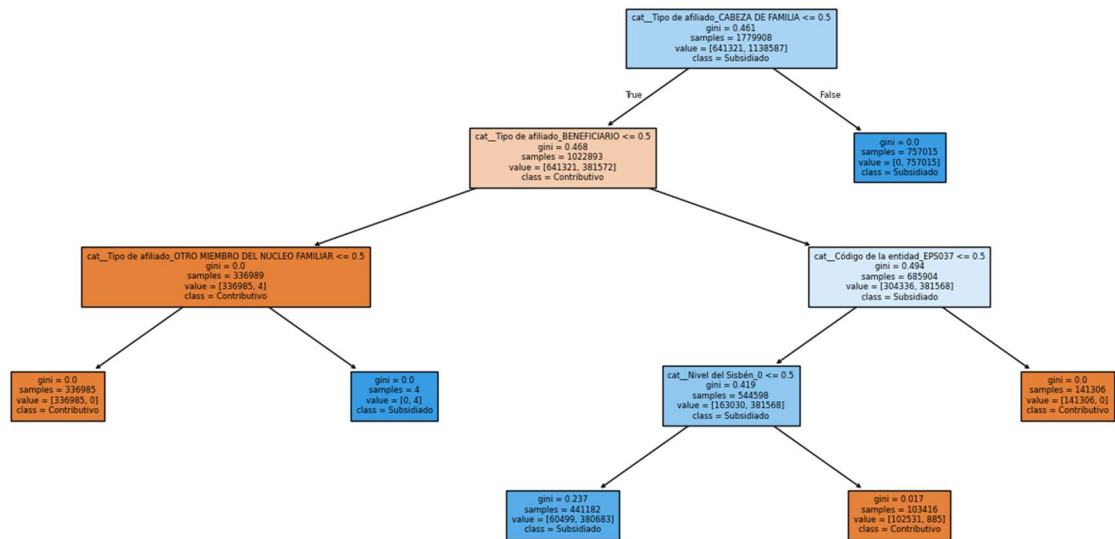
4. Análisis multivariado (SVD sobre One-Hot sparse)

La proyección multivariada mostró dos grupos de puntos diferenciados por color (régimen). Las bandas horizontales en el gráfico indican la existencia de estratos o grupos categóricos bien definidos, donde cada régimen se agrupa según combinaciones de variables.



- Se observa una clara separación entre los afiliados de régimen contributivo y subsidiado, lo que confirma que las variables analizadas contienen suficiente información para distinguir ambos grupos.
- Las zonas de superposición entre colores reflejan perfiles mixtos, por ejemplo, afiliados en transición o entidades que gestionan ambos regímenes.
- En conjunto, el análisis evidencia que los regímenes se distribuyen de forma estructurada y no aleatoria.

5. Árbol de decisión explicativo



El modelo de árbol identificó las siguientes reglas de clasificación:

- Variable raíz:** Tipo de afiliado
Es el principal determinante del régimen.
 - Cabeza de familia* → Subsidiado
 - Cotizante / Miembro del núcleo familiar* → Contributivo
- Segunda variable clave:** Código de la entidad
Determina si la EPS pertenece al régimen subsidiado o contributivo.
 - EPS específicas (p. ej. EPS037)* → Contributivo
 - Otras EPS* → Subsidiado
- Tercer nivel:** Nivel del Sisbén
Reafirma la dimensión socioeconómica.
 - Niveles bajos (0, 1, 2)* → Subsidiado
 - Niveles altos o no aplicables* → Contributivo

El árbol logra una separación casi perfecta (gini cercano a 0 en la mayoría de las hojas), demostrando que con solo tres variables se puede explicar la mayor parte de la variabilidad del régimen.

6. Conclusiones generales

Aspecto	Resultado principal
Predominio del régimen	Subsidiado ($\approx 64\%$), indicando mayor población en condiciones de vulnerabilidad.
Variables más influyentes	Tipo de afiliado, Código/Nombre de la entidad y Nivel del Sisbén.
Dimensión socioeconómica	Determina la pertenencia al régimen con alta precisión.
Dimensión institucional	Las EPS se asocian fuertemente con un régimen específico.
Dimensión demográfica	Género, edad o departamento no influyen significativamente.

7. Recomendaciones

- Utilizar las tres variables clave para construir modelos predictivos o dashboards de control de afiliación, optimizando la clasificación automática.
- Realizar un análisis geográfico complementario (por municipio o departamento) para identificar territorios con mayor proporción de subsidiados.
- Incorporar variables adicionales de ingreso, ocupación o formalidad laboral para mejorar la comprensión del fenómeno.