

Diagnóstico socioeconómico y territorial de poblaciones vulnerables

Unidad de Producción de Información Estadística
Equipo de Clasificación Socioeconómica
Comunas Popular, Santa Cruz y Manrique

Fuente: Base de datos 202630239420 anonimizada del Sisbén
IV diciembre 2025



Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación

Contenido

Equipo de clasificación socioeconómica

- 1. Conceptos clave y justificación del alcance
- 2. Metodología: datos, variables, modelo y validación
- 3. Resultados: desempeño, condiciones asociadas y territorio
- 4. Fase 2



Conceptos clave

Hogar (unidad de análisis)

Una persona o grupo de personas, parientes o no, que ocupan la totalidad o parte de una vivienda; atienden necesidades básicas con cargo a un presupuesto común y generalmente comparten las comidas.

Jefe/a del hogar

Es la persona que los miembros del hogar identifican como jefe. Para ser considerado como tal por parte de los miembros del hogar, debe haber un reconocimiento, ya sea por razones económicas, de edad o de autoridad.

Convención «A/B»: es «A o B»

En todo el informe, la barra de «A/B» se lee como disyunción: hogares del grupo A o del grupo B, tomados como un solo conjunto de comparación. No es un cociente ni una fracción.

Grupos de clasificación del Sisbén IV

Grupo A — Pobreza extrema. 5.850 hogares en el universo modelado (6,2 %).

Grupo B — Pobreza moderada. 25.608 hogares (27,3 %).

Grupo C — Vulnerabilidad: hogares no pobres pero en riesgo de caer en pobreza. 62.176 hogares (66,4 %).

Grupo D — No pobre, no vulnerable. 32.096 hogares, excluidos del modelo porque el objetivo es discriminar dentro de la población pobre y vulnerable.

Riesgo (score del hogar)

Probabilidad estimada (entre 0 y 1) de que el hogar presente un perfil propio de los grupos A o B, es decir, mide similitud de perfil A/B dentro de los hogares clasificados en C.



LASSO y métricas del modelo

LASSO — Least Absolute Shrinkage and Selection Operator

En español «operador de contracción y selección por mínimos absolutos») es un método de regresión con penalización L1 que suma el valor absoluto de los coeficientes. Cumple dos funciones simultáneas:

- (i) Estima el riesgo del hogar
- (ii) selecciona variables automáticamente, llevando exactamente a cero los coeficientes de variables poco informativas o redundantes.

Variable de resultado del modelo:

$y = 1$ si el hogar es A o B

$y = 0$ si es C

AUC (Area Under the Curve)

Mide qué tan bien el modelo distingue entre hogares A/B y hogares C. Un AUC más alto indica mayor capacidad para ordenar correctamente los hogares según su riesgo.

- 1.0:** Discriminación perfecta.
- 0.5:** Capacidad predictiva equivalente al azar.
- < 0.5:** Discriminación inversa (predicción sistemáticamente invertida).

En el caso de estudio: probabilidad de que un hogar A/B elegido al azar reciba mayor score que un hogar C al azar.

Puntaje de Brier (Brier Score)

Evalúa qué tan cercanas son las probabilidades estimadas por el modelo frente a los resultados observados. Mientras más bajo sea el valor, mejor es la calidad de la predicción probabilística.

OOF (Out-Of-Fold) : fuera de muestra

Son predicciones fuera de muestra generadas mediante validación cruzada. Permiten evaluar el desempeño del modelo en datos que no fueron usados para entrenarlo.

Odds ratio (OR) : Razón posibilidades

Mide cómo cambia la posibilidad de pertenecer al perfil vulnerable cuando una característica está presente.

- OR = 1:** No existe asociación entre la variable y el evento.
- OR > 1:** La presencia de la variable incrementa los odds (posibilidades) de que ocurra el evento.
- OR < 1:** La presencia de la variable disminuye los odds (posibilidades) de que ocurra el evento.



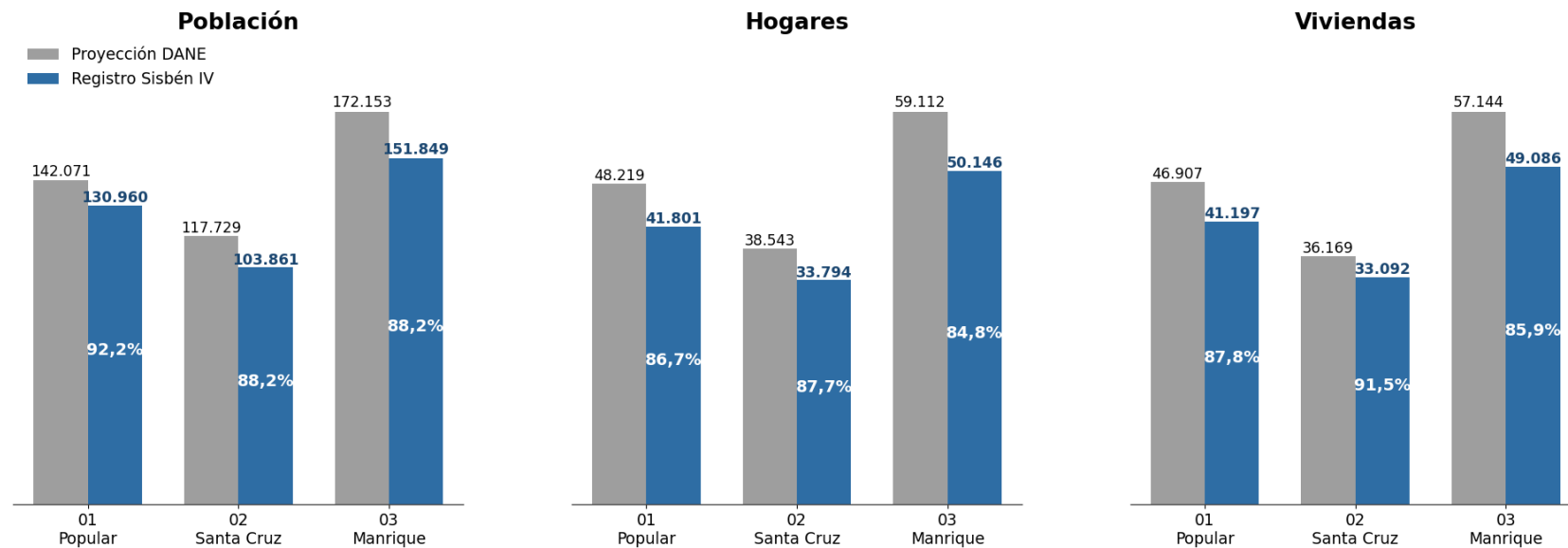
Justificación selección de comunas Popular, Santa Cruz y Manrique

Cobertura del Sisbén IV frente a las proyecciones del DANE (2025)

El Sisbén IV cubre entre el 84,8 % y el 92,2 % de la población, los hogares y las viviendas que el DANE proyecta para las tres comunas: una base administrativa con alta cobertura para caracterizarlas. Al ser un registro de demanda, no incluye a los hogares que no han solicitado la encuesta.

Cobertura del Sisbén IV frente a las proyecciones del DANE — Comunas 1, 2 y 3

Personas, hogares y viviendas registrados en el Sisbén IV (dic. 2025) como % de lo proyectado por el DANE · el porcentaje dentro de la barra azul es la cobertura



Fuente: DANE — proyecciones de población, hogares y viviendas, contrato interadministrativo 4600104690 de 2025 (act. 30/07/2025 y 28/11/2025); base anonimizada Sisbén IV, dic. 2025.

Gráfico. IMCV por comunas y corregimientos 2024 - 2025, Distrito de Medellín, 2025

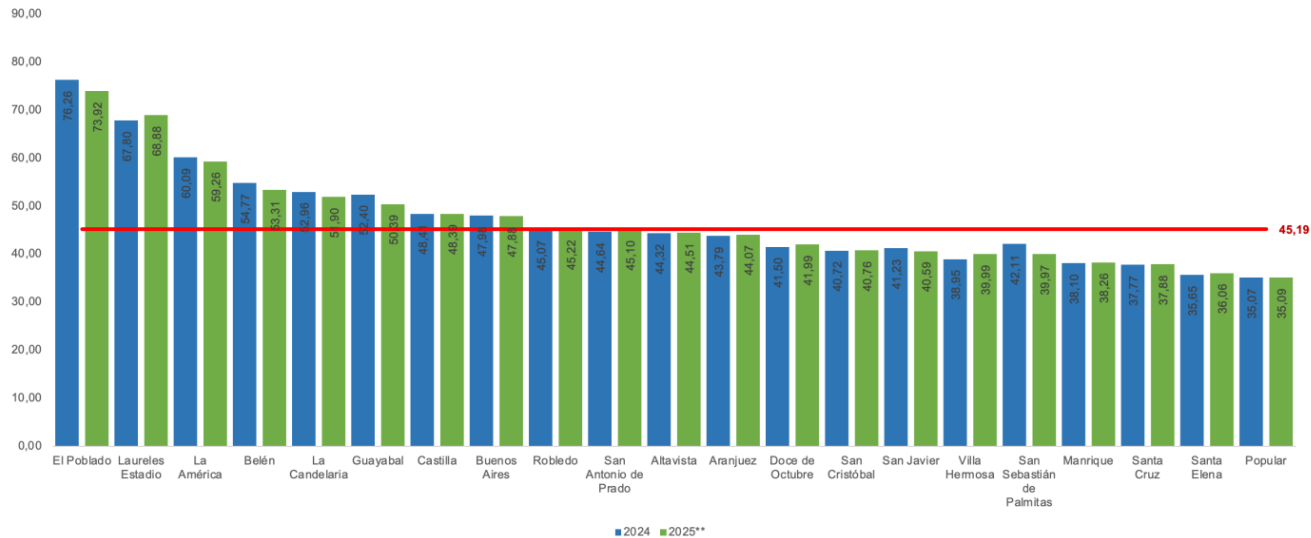
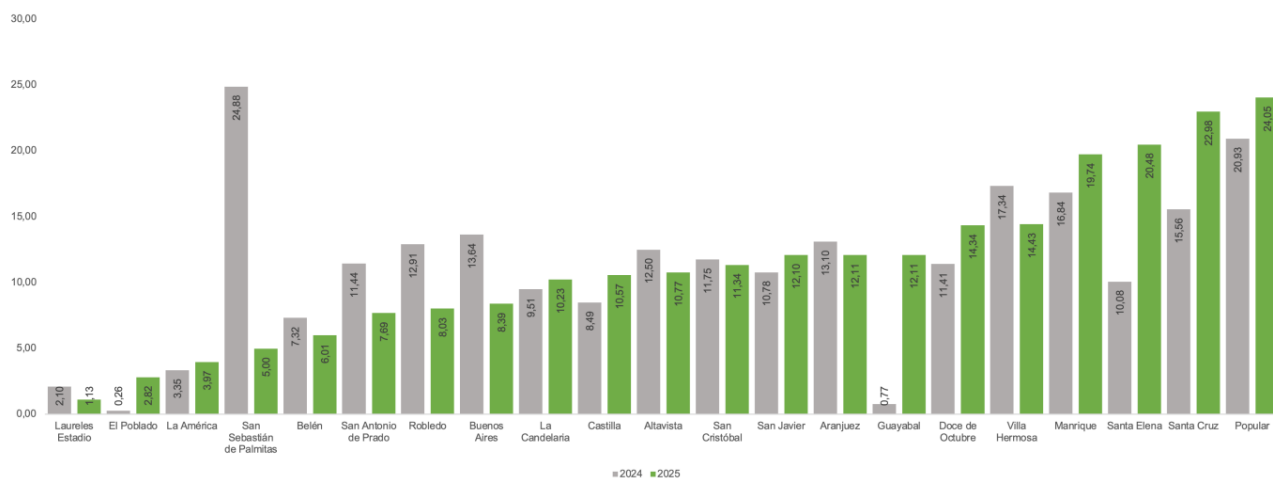
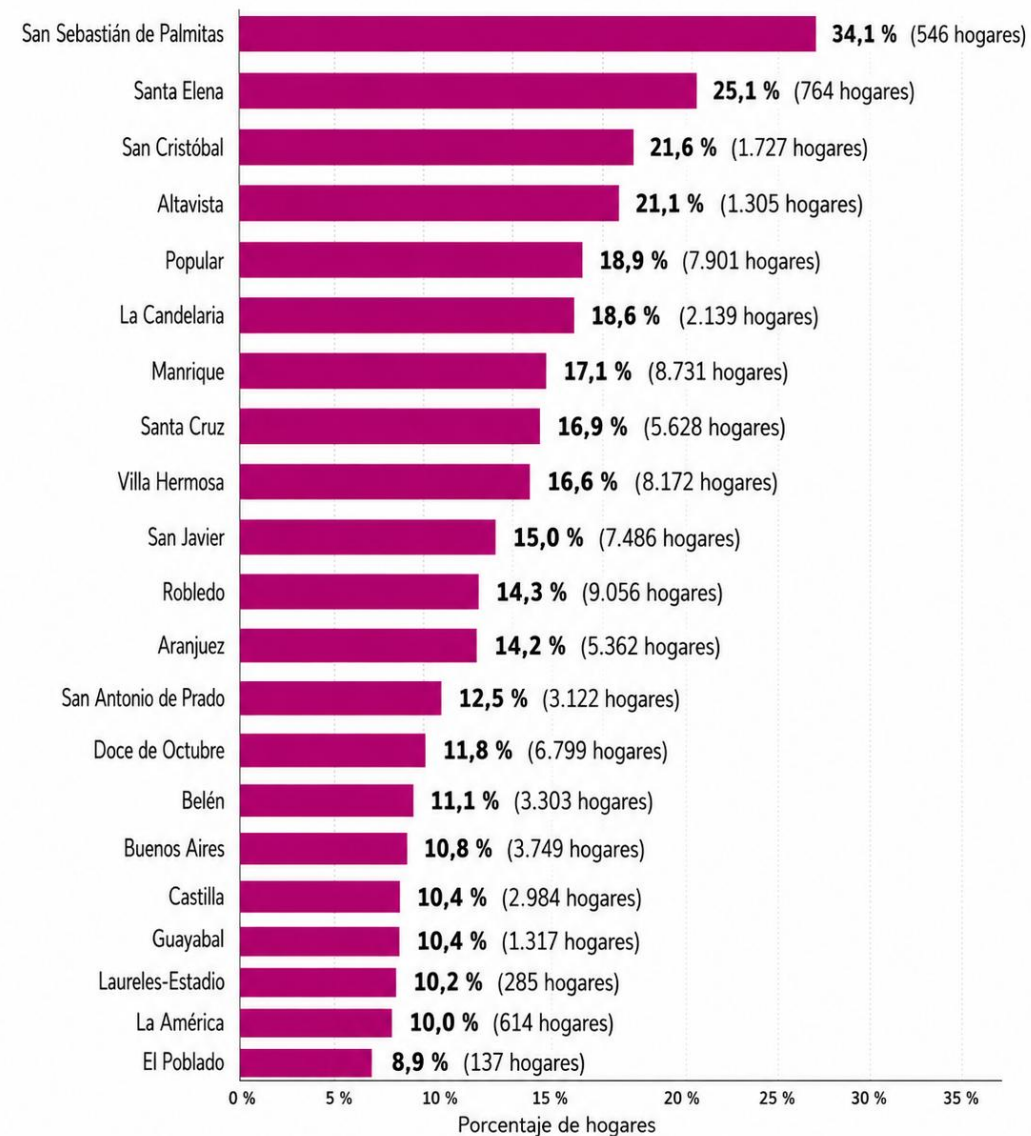


Gráfico. IPM por comunas y corregimientos 2024 - 2025, Distrito de Medellín, 2025



Incidencia de la pobreza multidimensional

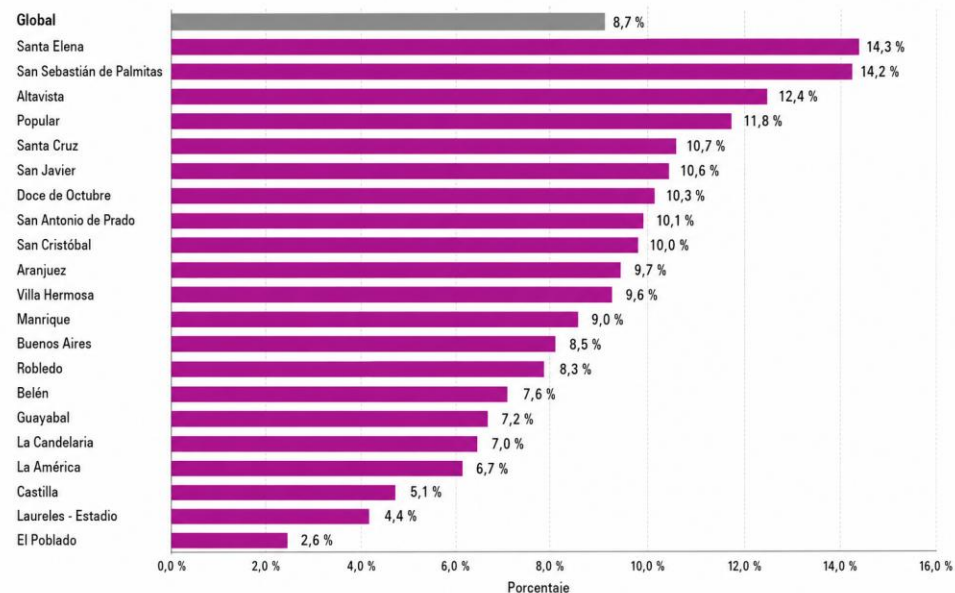
Sisbén IV 2025 · Porcentaje de hogares en pobreza multidimensional sobre el total de hogares registrados



Fuente: Alcaldía de Medellín. (2026). Caracterización Socioeconómica Sisbén IV. Departamento Administrativo de Planeación: Subdirección de Prospectiva, Información y Evaluación Estratégica.

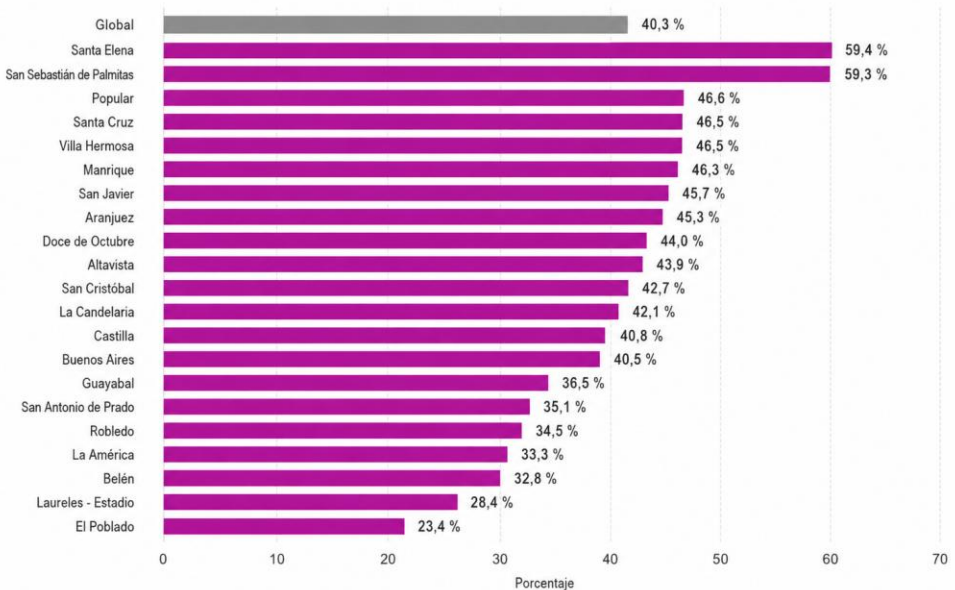
Tasa de desempleo 2024

Comunas, corregimientos y valor global de Medellín



Tasa de informalidad 2024

Comunas, corregimientos y valor global de Medellín



Subdirección de Prospectiva, Información y Evaluación Estratégica

Departamento Administrativo de Planeación



Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación

Promedio de inversión en PP por comuna y corregimiento

Tabla N° 2

2013 - 2024

Comuna / Corregimiento	Promedio	Comuna / Corregimiento	Promedio
Popular	7.12%	San Antonio de Prado	4.50%
Manrique	6.37%	Guayabal	4.11%
Santa Cruz	6.27%	Santa Elena	3.89%
Doce de Octubre	6.25%	Candelaria	3.86%
Robledo	6.23%	La América	3.82%
San Javier	6.05%	San Cristóbal	3.62%
Villa Hermosa	6.02%	San Sebastián de Palmitas	3.00%
Aranjuez	5.64%	Laureles/Estadio	2.97%
Castilla	5.00%	Altavista	2.81%
Buenos Aires	4.99%	Poblado	2.81%
Belén	4.68%		

■ Por encima del promedio general

■ Por debajo del promedio general

Promedio general: 4.76%

Fuente: Tabla N° 2 — Promedio de inversión en Presupuesto Participativo por comuna y corregimiento (2013-2024)

Las tres comunas superan el agregado de ciudad en desocupación e informalidad (2024) y han sido prioridad histórica en la asignación territorial de recursos del PP.

Fuentes: tasas de desocupación e informalidad 2024 — Indicadores y estadísticas del Distrito, Departamento Administrativo de Planeación [tablero de Power BI], Alcaldía de Medellín. Presupuesto Participativo — OPPCM (2024),



VARIABLES DEL MODELO: DEFINICIÓN, PREGUNTA Y CONSTRUCCIÓN (1 DE 2)

Variable	Definición operativa	Pregunta de la ficha	Cómo se construye	Variables de la base
1. Composición y carga demográfica del hogar — estructura del hogar y carga potencial de cuidado o sostenimiento económico				
Razón de dependencia	Proporción de miembros dependientes sobre el total del hogar	P.51	(menores de 18 + personas de 60 o más) / tamaño del hogar	fec_nacimiento, tamaño del hogar
Tamaño del hogar	Número de personas que componen el hogar	—	Conteo de miembros por identificador de hogar	ide_hogar, registro
Hogar monoparental con menores de 18	Jefe/a sin cónyuge conviviente y con al menos un menor de 18	P.55, P.56, P.51	= 1 si el jefe/a no tiene cónyuge en el hogar y hay menores de 18	tip_estado_civil, ind_conyugue_vive_hogar, fec_nacimiento, tip_parentesco
2. Capital humano y educación — condiciones educativas del jefe/a y de los miembros del hogar				
Años de educación del jefe/a	Años de educación del jefe/a	P.70	Conversión de nivel + grado a años	niv_educativo, grado_alcanzado, tip_parentesco
Analfabetismo en el hogar	Al menos un miembro de 15+ que no sabe leer ni escribir	P.68, P.51	= 1 si algún miembro de 15 o más años responde «no» a saber leer y escribir	ind_leer_escribir, fec_nacimiento
Inasistencia escolar 6 a 16	Al menos un miembro de 6 a 16 años que no estudia	P.69, P.51	= 1 si algún miembro de 6 a 16 años no asiste a un establecimiento educativo	ind_estudia, fec_nacimiento
3. Inserción laboral del jefe/a — situación ocupacional del jefe/a y posible fragilidad en la generación de ingresos				
Jefe/a buscando trabajo	Jefe/a desempleado/a que busca trabajo activamente	P.72 (op. 2)	= 1 si la actividad del mes del jefe/a es «buscando trabajo»	tip_actividad_mes, tip_parentesco
Jefe/a sin actividad	Corresponde a las personas cuya actividad no se ajusta a ninguna de las opciones anteriores.	P.72 (op. 0)	= 1 si la actividad del mes del jefe/a es «sin actividad»	tip_actividad_mes, tip_parentesco
Jefe/a en oficios del hogar	Jefe/a dedicado/a a oficios del hogar no remunerados	P.72 (op. 4)	= 1 si la actividad del mes del jefe/a es «oficios de hogar»	tip_actividad_mes, tip_parentesco
4. Cargas adicionales de cuidado — presencia de personas con mayores necesidades de cuidado o sin protección suficiente				
Vejez desprotegida	Al menos un mayor de 60 sin ninguna protección a la vejez	P.51, P.71, P.72, P.77	= 1 si ≥ 1 miembro de 60+ no cotiza, no es pensionado, no recibe pensión ni Colombia Mayor	ind_fondo_pensiones, ind_ingr_pension, vlr_ingr_col_mayor
Discapacidad en el hogar	Al menos un miembro con limitación permanente	P.59	= 1 si algún miembro reporta alguna de las siete limitaciones permanentes	ind_discap_* 7 (ver, oír, hablar, moverse, bañarse, vestirse, salir a la calle sin ayuda, aprender)



VARIABLES DEL MODELO: DEFINICIÓN, PREGUNTA Y CONSTRUCCIÓN (2 DE 2)

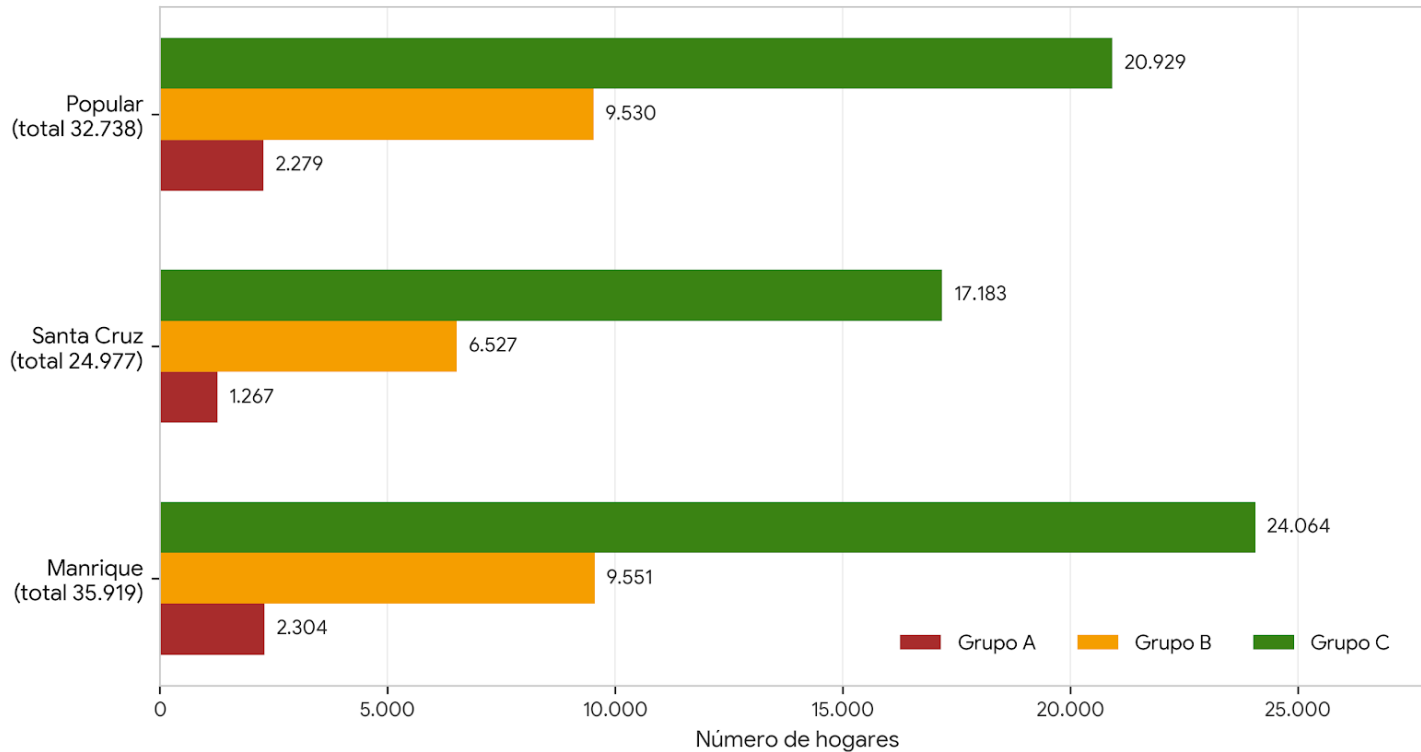
Variable	Definición operativa	Pregunta de la ficha	Cómo se construye	Variables de la base
5. Capacidad económica y activos del hogar — capacidad de consumo y acumulación básica de bienes				
Gasto per cápita (log)	gasto mensual del hogar	P.41	$\ln(1 + \text{gasto total del hogar} / \text{tamaño del hogar})$	vlr_gasto_* (8 rubros), tamaño del hogar
Número de activos	Conteo de bienes durables del hogar (0–8)	P.40	Suma de ocho indicadores de tenencia de bienes durables	ind_tiene_nevera, _lavadora, _pc, _internet, _moto, _tractor, _carro, _bien_raiz
6. Condiciones habitacionales y tenencia — calidad de la vivienda, estabilidad residencial y presión habitacional				
Vivienda inadecuada	Vivienda con paredes o piso de materiales precarios	P.18, P.19	= 1 sin paredes/guadua/caña/esterilla/otro vegetal / zinc/tela/lona/cartón/latas/desechos/ plástico o piso en tierra o arena	tip_mat_paredes, tip_mat_pisos
Tenencia en arriendo	Hogar que paga arriendo o subarriendo	P.24 (op. 1)	= 1 si la forma de ocupación de la vivienda es «arriendo o subarriendo»	tip_ocupa_vivienda
Tenencia ocupante de hecho	Hogar en posesión sin título de la vivienda	P.24 (op. 5)	= 1 si la forma de ocupación es «posesión sin título, ocupante de hecho»	tip_ocupa_vivienda
Hacinamiento crítico	Tres o más personas por cuarto exclusivo para dormir	P.27	= 1 si tamaño / cuartos exclusivos ≥ 3 , o si no hay cuarto exclusivo para dormir	num_cuartos_unicos_dormir, tamaño del hogar
7. Servicios básicos y riesgo territorial — acceso a condiciones básicas de saneamiento y exposición a eventos físicos ambientales				
Saneamiento inadecuado	Sin alcantarillado o servicio sanitario precario	P.20, P.28	= 1 si no hay alcantarillado, o el sanitario es sin conexión/letrina/sin servicio (cód. 3–5)	ind_tiene_alcantarillado, tip_sanitario
Afectación por eventos naturales	Hogar afectado por un evento de riesgo ambiental	P.43	= 1 si declara inundación, avalancha, terremoto, incendio, vendaval o hundimiento	ind_evento_* (6 columnas)



Hogares en el modelo por grupo y comuna

Hogares incluidos en el modelo por grupo y comuna

Grupos A, B y C con información completa · Popular, Santa Cruz y Manrique · total 93.634 hogares



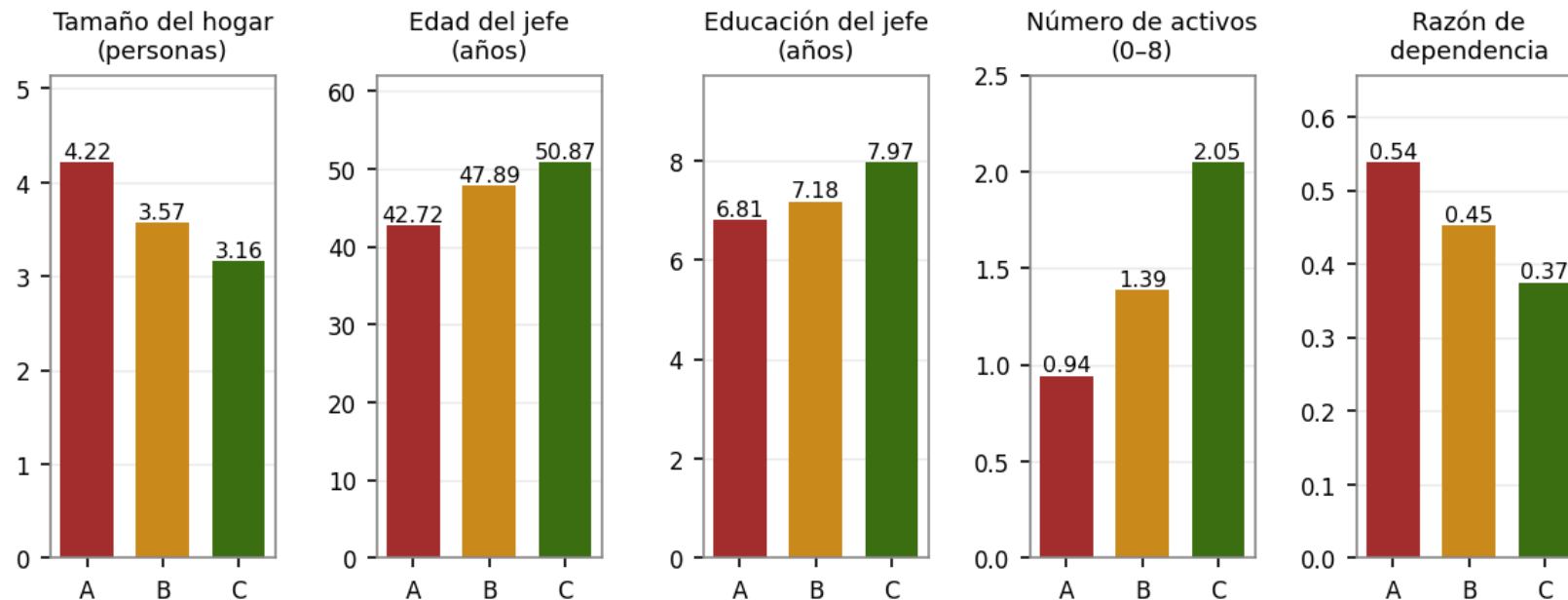
El universo del modelo son 93.634 hogares A, B o C. El grupo C concentra el 66,4 %, seguido de B (27,3 %) y A (6,2 %).



Promedios de las variables continuas, según grupo

Promedios por hogar de las variables continuas, según grupo Sisbén IV

El gradiente es monótono de C a A en las cinco variables



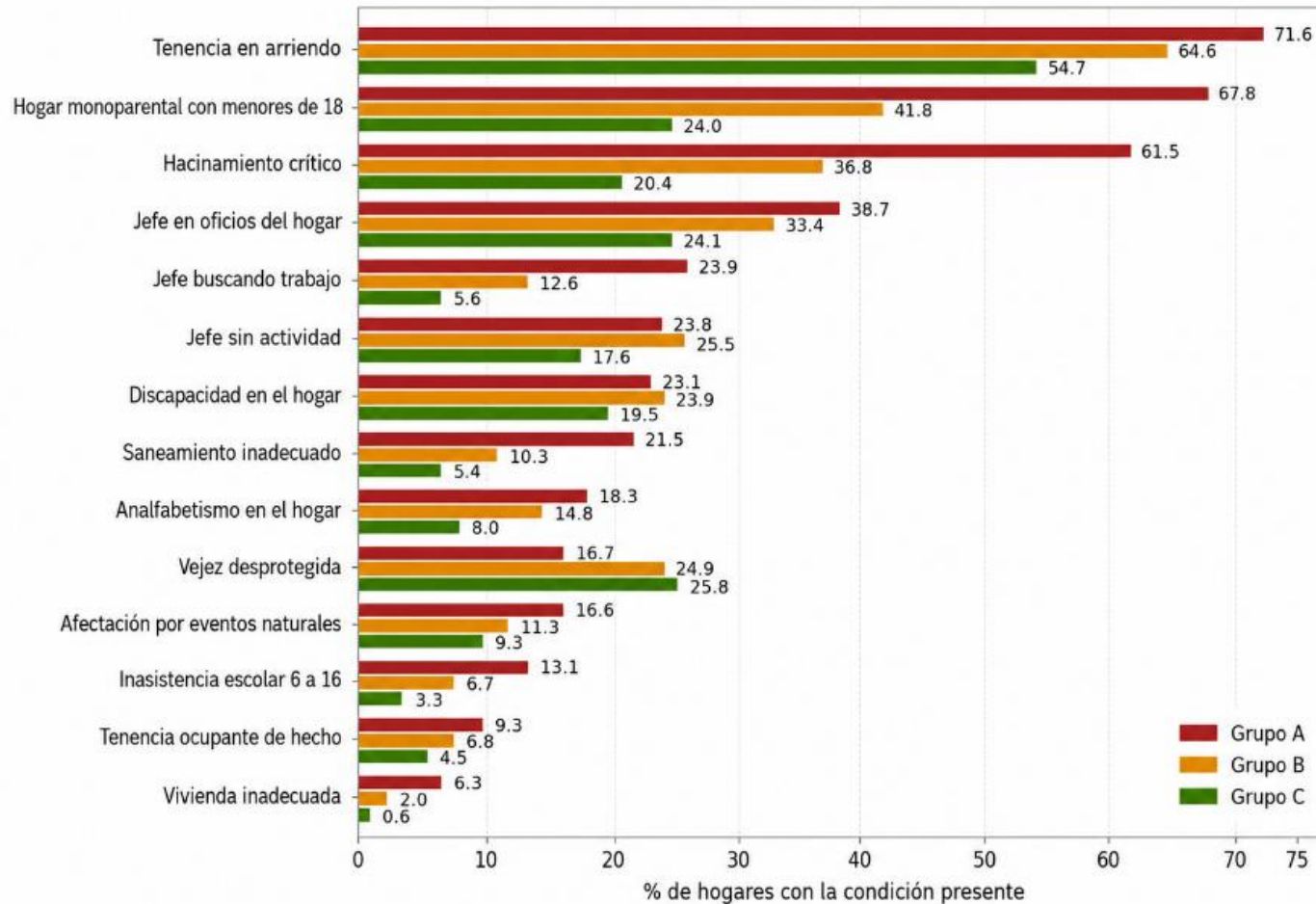
Las variables construidas reproducen el gradiente esperado C → A. Al pasar de C a A los hogares son más grandes (3,2 → 4,2 personas), con jefes/as más jóvenes (51 → 43 años), menos educación (8,0 → 6,8 años), menos activos (2,1 → 0,9 de 8) y mayor razón de dependencia (0,37 → 0,54). El gradiente es monótono en las cinco variables, lo que respalda su validez de constructo.



Prevalencia de las condiciones binarias, según grupo

Prevalencia de las condiciones binarias del modelo, según grupo

Comunas 1, 2 y 3 — 2025



El mismo gradiente aparece en las privaciones.

Las brechas más marcadas: monoparentalidad con menores (67,8 % en A vs. 24,0 % en C), desempleo del jefe/a (25,9 % vs. 5,6 %) y vivienda inadecuada (6,2 % vs. 0,6 %).

Una condición rompen el patrón —vejez desprotegida (más alta en C) —, coherente con la demografía de cada grupo.



Métricas de desempeño del modelo LASSO (A/B vs. C)

Métrica	Valor	Detalle
AUC fuera de muestra (OOF)	0,8365	IC 95 %: [0,8339 – 0,8392] · bootstrap percentil, B = 1.999
AUC dentro de muestra	0,8367	Brecha con OOF: 0,0002 (sin sobreajuste relevante)
Brier fuera de muestra	0,1522	Error cuadrático medio de las probabilidades
Brier dentro de muestra	0,1521	—

El AUC de 0,8365 muestra que el modelo discrimina bien entre los hogares A/B y C. Al estar bastante por encima de 0,5, el modelo no está clasificando al azar; y al acercarse a 1, muestra una buena capacidad para asignar scores más altos a los hogares con perfil A/B.

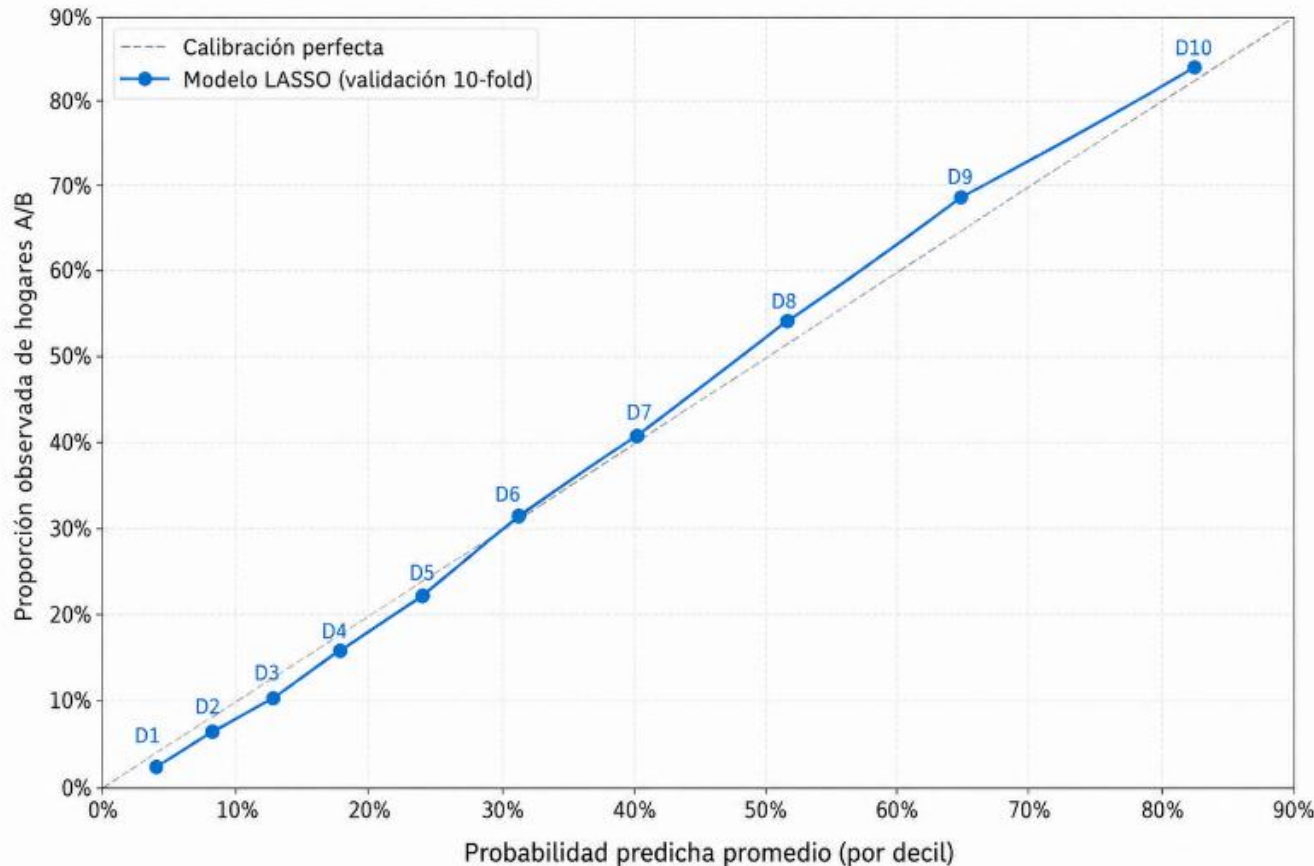
El intervalo de confianza del AUC se estimó mediante bootstrap percentil con 1.999 remuestras, lo que permite evaluar la estabilidad del desempeño del modelo. En cada remuestra se recalcula el AUC y luego se toman los percentiles 2,5 y 97,5 para obtener el intervalo al 95%



Calibración por deciles del score (fuera de muestra)

Calibración del modelo por decil de riesgo

El modelo separa los deciles y sus probabilidades coinciden con las frecuencias observadas (OOF)



Los puntos siguen de cerca la diagonal de calibración perfecta: la probabilidad predicha se aproxima con la frecuencia real observada.

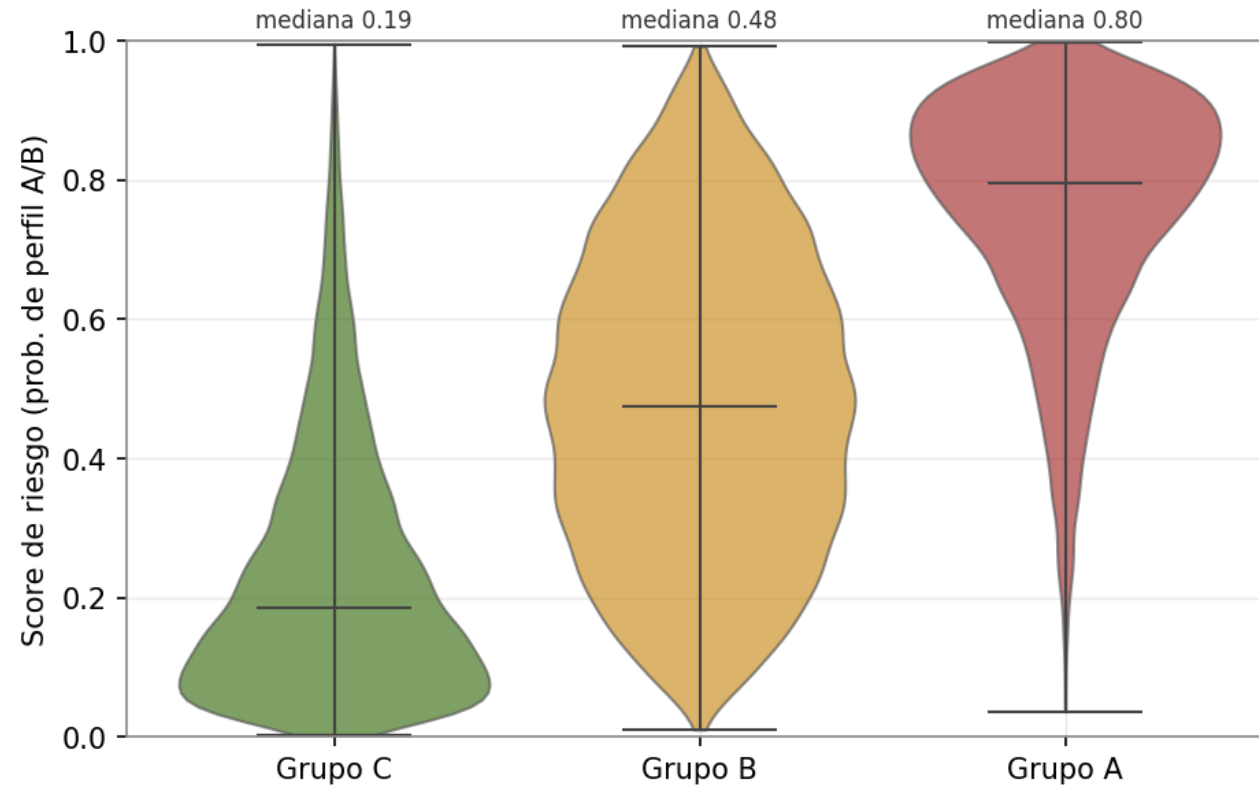
En los hogares de mayor score el riesgo está, si acaso, ligeramente subestimado.



Distribución del score de riesgo según grupo

Distribución del score de riesgo según grupo Sisbén IV

La cola superior del grupo C es la población objetivo del modelo



La cola alta del grupo C es la población objetivo.

La mediana del score crece de 0,19 en C a 0,48 en B y 0,80 en A. La separación entre A y C es nítida; B se solapa con ambos.

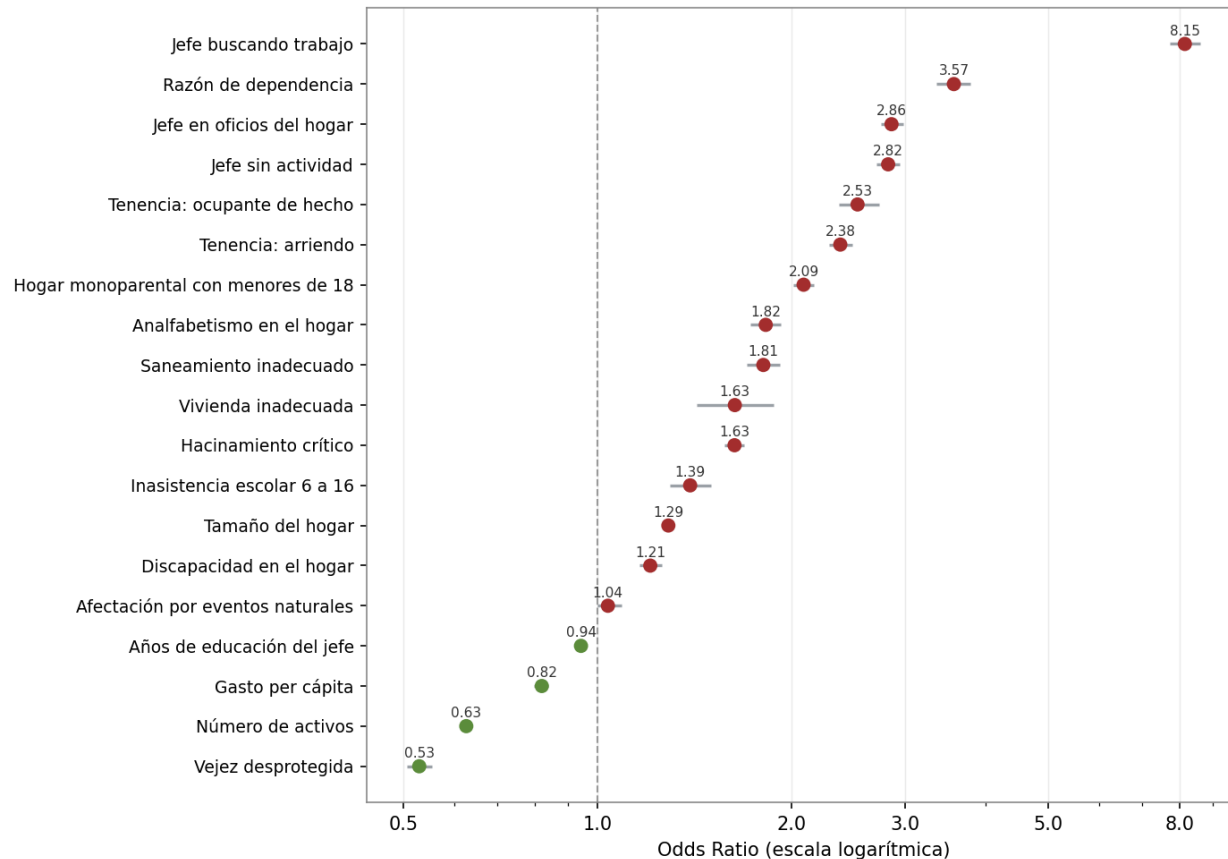
La cola superior del grupo C (parte alta del violín verde) son hogares clasificados C cuyo score los sitúa en el rango típico de A/B: ese es, exactamente, el blanco del modelo.



Odds ratios del modelo con IC 95 % (escala logarítmica)

Condiciones asociadas al perfil de pobreza — Odds Ratios con IC 95 %

Verde: reduce la probabilidad de perfil A/B (OR<1) · Rojo: la aumenta (OR>1) · bootstrap, B=1.999



La condición más asociada al perfil de pobreza es el desempleo del jefe/a (OR = 8,2), seguida de la razón de dependencia (OR = 3,6), la inactividad u oficios del hogar del jefe/a (OR ≈ 2,8–2,9) y la tenencia (OR = 2,4–2,5).

En sentido protector operan los activos, la educación del jefe/a y el gasto per cápita.

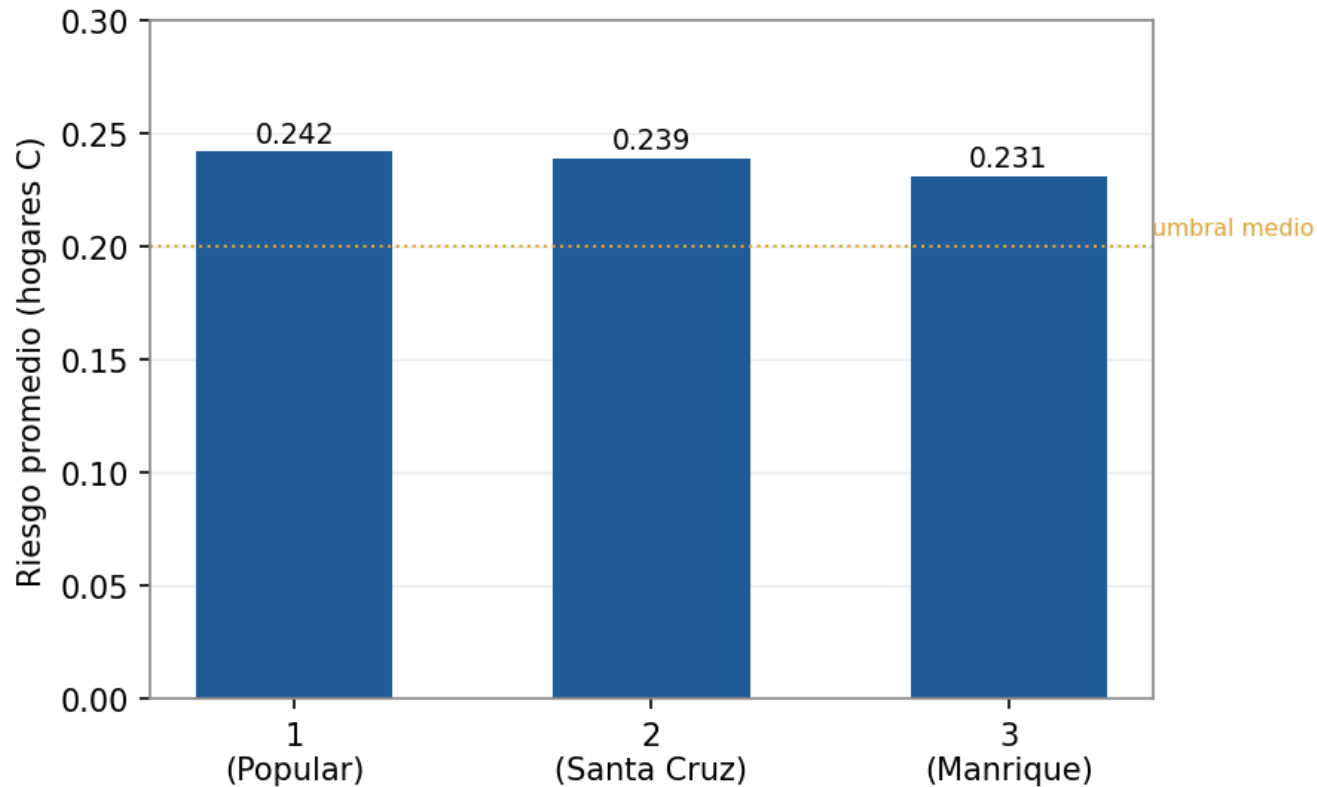
La vejez desprotegida aparece con OR = 0,53 debido a la parte demográfica.



Riesgo promedio de los hogares C por comuna

Riesgo promedio de los hogares del grupo C por comuna

La comuna es una escala demasiado gruesa para focalizar

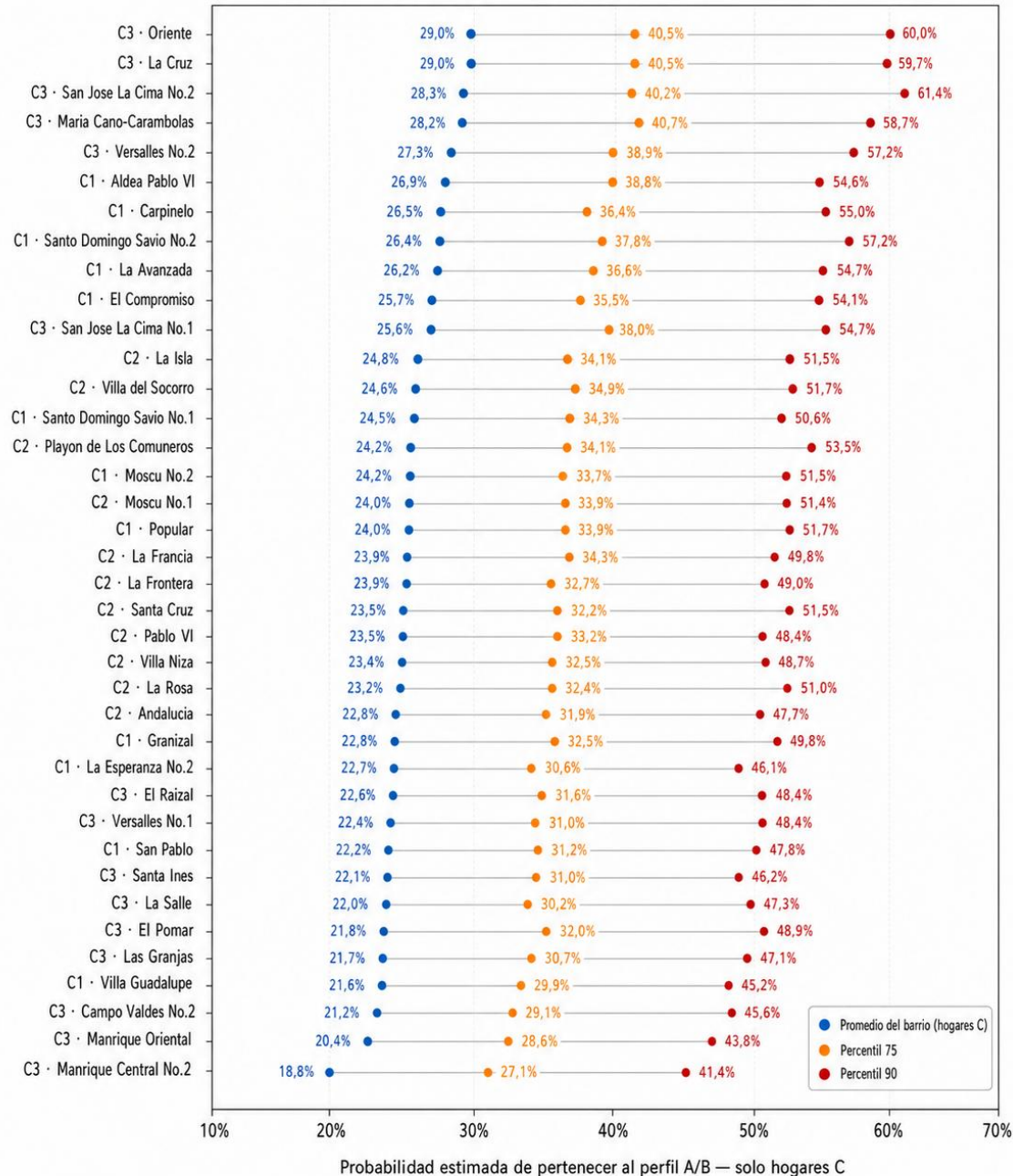


El riesgo promedio de los hogares C es casi idéntico en las tres comunas: 0,242 (Popular), 0,239 (Santa Cruz) y 0,231 (Manrique).

La comuna es una escala demasiado gruesa para priorizar.

Heterogeneidad interna del riesgo — 38 barrios (solo hogares C)

Un barrio con promedio moderado puede contener hogares con riesgo extremo (p90 ≈ 50%)



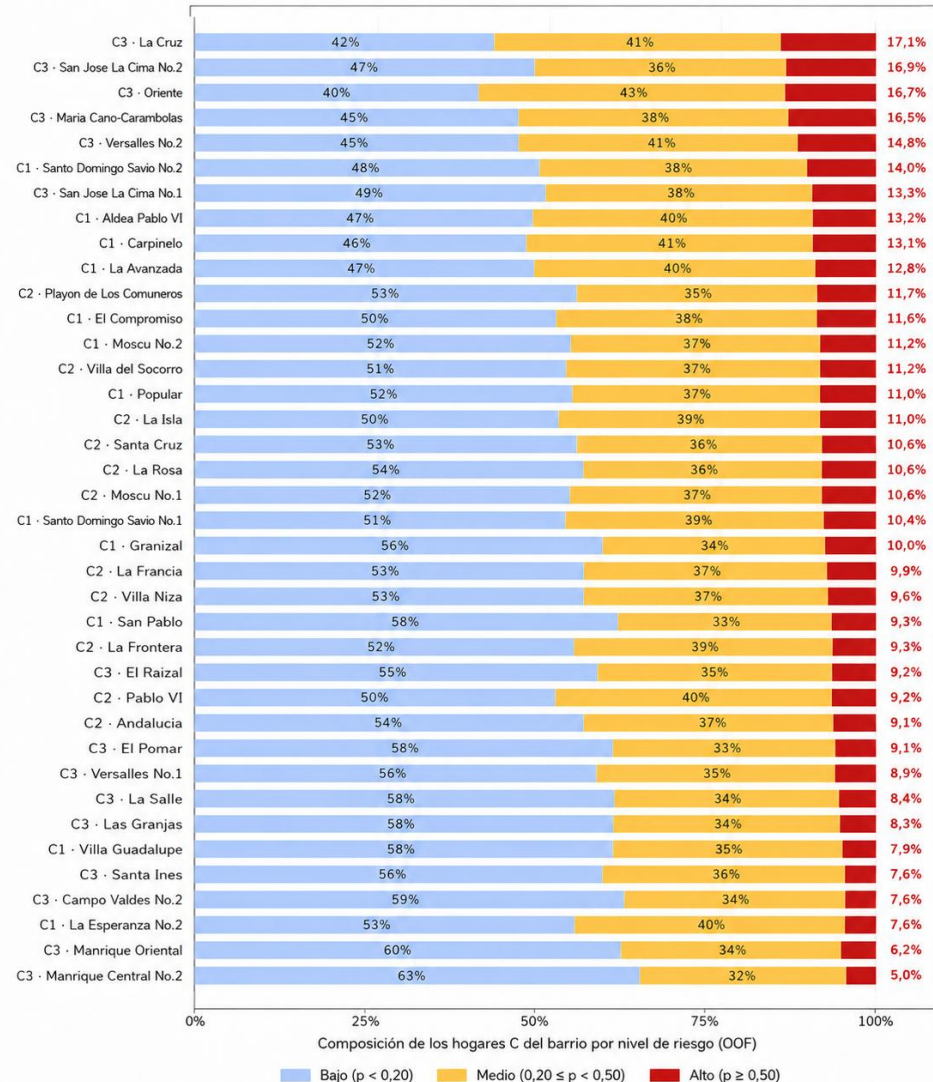
Heterogeneidad interna del riesgo

Aunque el promedio de riesgo de un barrio puede parecer moderado, al interior existen hogares con niveles de riesgo mucho más altos. La comparación entre el promedio del barrio, el percentil 75 y el percentil 90 muestra que algunos hogares dentro del mismo territorio concentran una probabilidad estimada considerablemente mayor. Por eso, el análisis no debe quedarse solo en el promedio barrial, sino identificar los hogares con mayor riesgo dentro de cada barrio.



Composición de cada barrio por nivel de riesgo – hogares del grupo C

Niveles anclados en la probabilidad calibrada: bajo ($p < 0,20$), medio ($0,20 \leq p < 0,50$), alto ($p \geq 0,50$) - ordenado por % en nivel alto (cifra roja)



Composición de cada barrio por nivel de riesgo

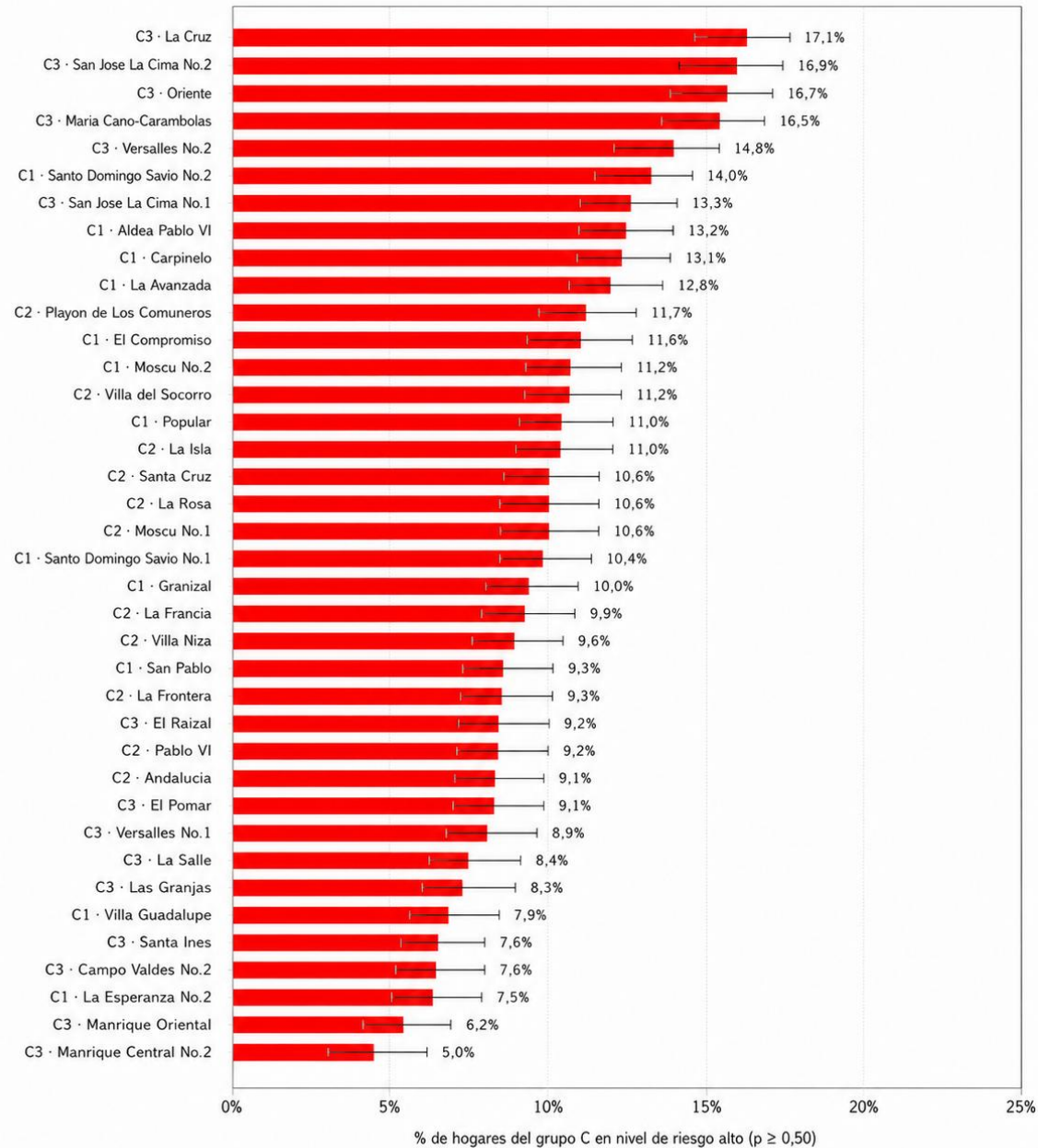
Lo que ordena el territorio es la franja roja (nivel alto).

En todos los barrios el nivel bajo agrupa entre 40 % y 63 % de los hogares C y el nivel medio entre 32 % y 43 %.

La diferencia que jerarquiza el territorio es la franja roja del nivel alto. Con el umbral operativo ($p \geq 0,50$), 6.423 hogares C —el 10,3 % de los 62.176— presentan perfil de pobreza.

Hogares del grupo C en nivel de riesgo alto, por barrio

Barras: % de hogares C con $p \geq 0,50$ (OOF) · líneas: IC 95% de Wilson



Porcentaje de hogares C en nivel alto, por barrio

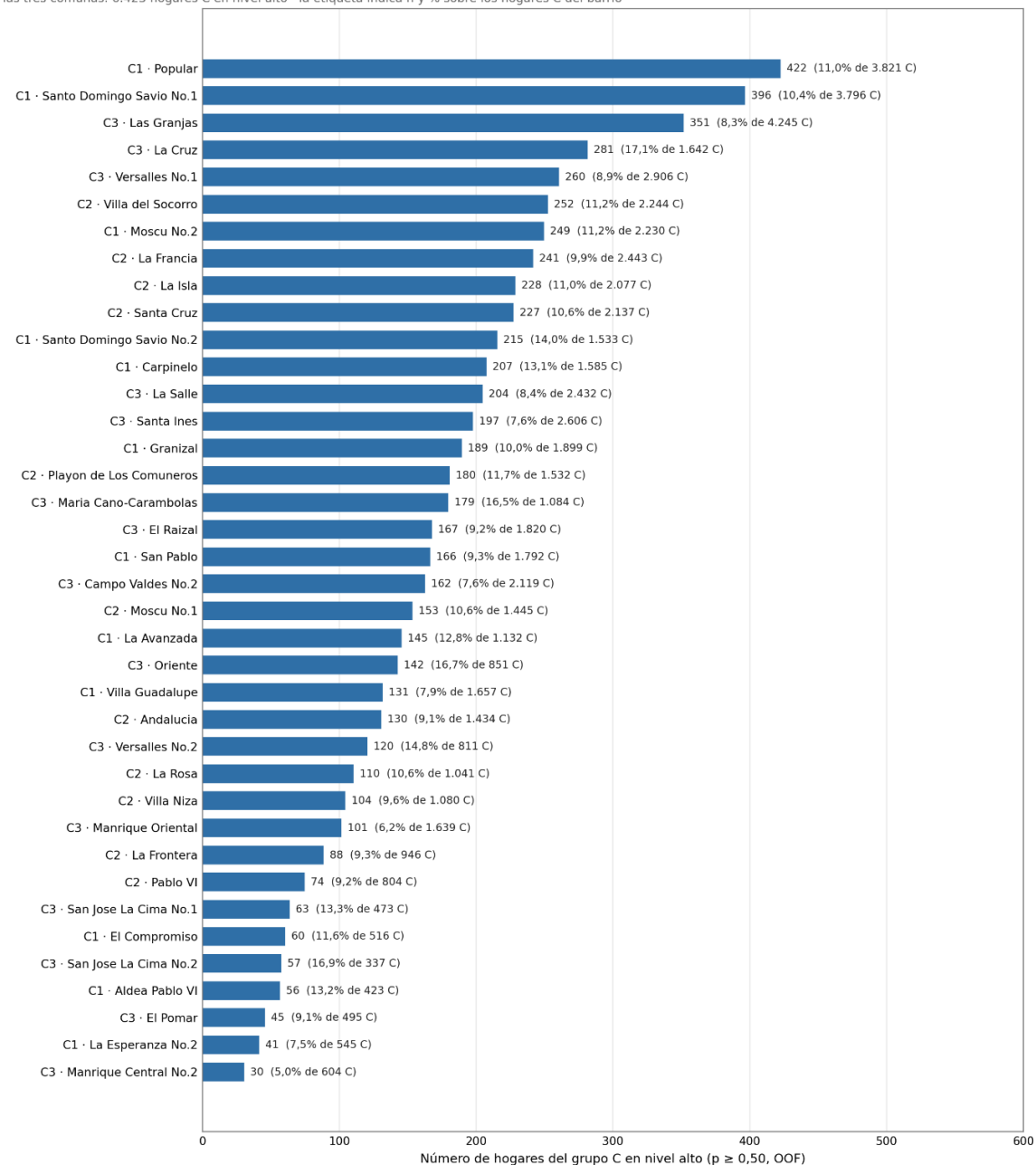
Incidencia: dónde se concentran los casos críticos.

El porcentaje de hogares C en nivel alto ($p \geq 0,50$) va de 5,0 % en Manrique Central No.2 a 17,1 % en La Cruz, con IC 95 % de Wilson.

Encabezan la ladera alta de las comunas 3 y 1. Esta es la lectura de incidencia: indica dónde la densidad de casos críticos permite una intervención rápida y de alto rendimiento.

Hogares del grupo C en nivel alto por barrio — magnitud absoluta

Total en las tres comunas: 6.423 hogares C en nivel alto · la etiqueta indica n y % sobre los hogares C del barrio



ión de Prospectiva, Información y Evaluación Estratégica

Departamento Administrativo de Planeación



Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación

Número de hogares C en nivel alto, por barrio

Volumen: dónde están en mayor número.

En magnitud absoluta —la que dimensiona la operación— los mayores volúmenes están en los barrios grandes: Popular (422), Santo Domingo Savio No.1 (396), Las Granjas (351), La Cruz (281) y Versalles No.1 (260).

La Cruz combina alto porcentaje (17,1 %) y alto volumen: es el punto de entrada operativo natural. La estrategia eficiente lee incidencia y volumen a la vez.

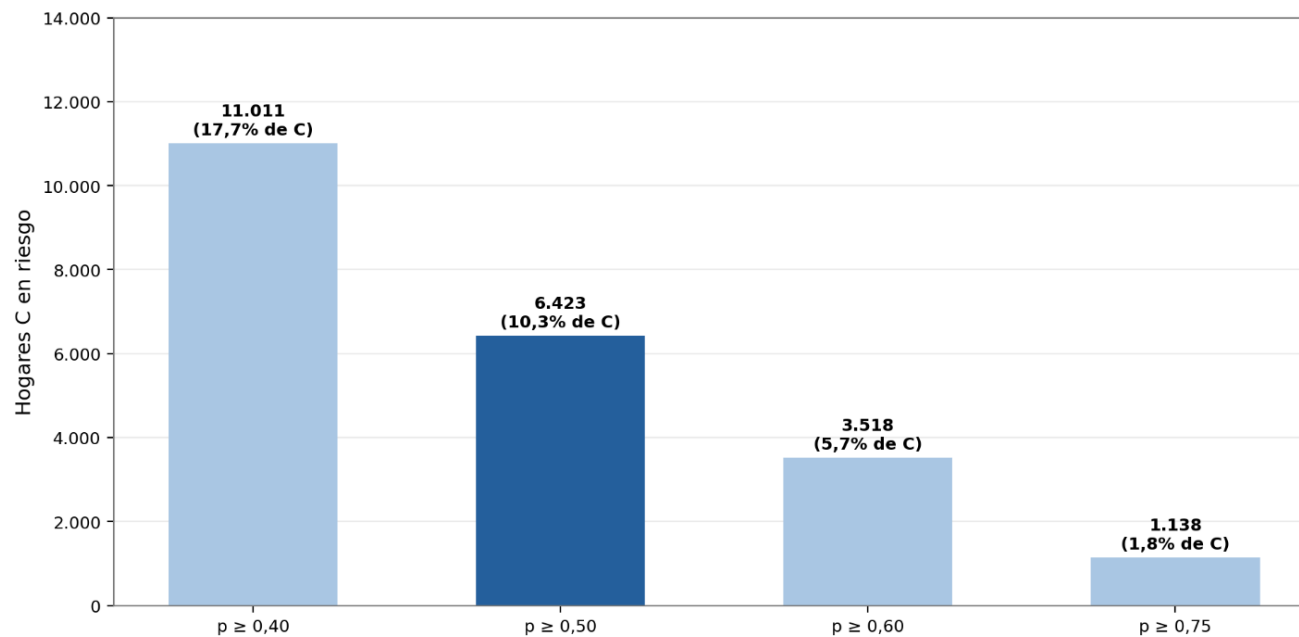
Fuente: base anonimizada Sisbén IV, dic. 2025.



El tamaño de la población objetivo es una decisión de umbral

Sensibilidad al umbral: hogares C en riesgo según el corte elegido

Predicciones fuera de muestra · Umbral operativo $p \geq 0,50$



Umbral	Hogares C en riesgo	% del grupo C
$p \geq 0,40$	11.011	17,7 %
$p \geq 0,50$ (operativo)	6.423	10,3 %
$p \geq 0,60$	3.518	5,7 %
$p \geq 0,75$	1.138	1,8 %

Total de hogares C con score: 62.176.

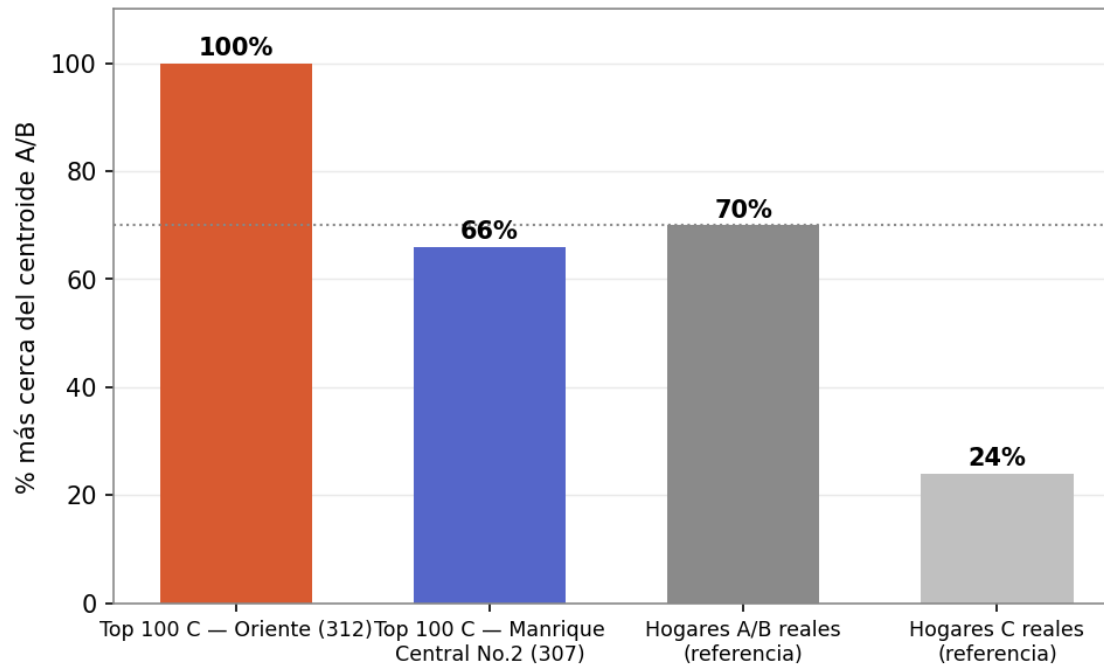
A mayor exigencia del umbral, menor población objetivo y mayor certeza de que cada hogar señalado comparte el perfil de pobreza. La elección depende de la capacidad operativa y del costo de los errores de inclusión y exclusión.



Validación de perfil: cercanía al centroide A/B

Validación de perfil: porcentaje de hogares más cercanos al centroide A/B

Los hogares de mayor score se parecen, en su perfil completo, a los hogares en pobreza



Los hogares señalados se parecen a hogares pobres.

Entre los 100 hogares C de mayor score de Oriente (mayor riesgo), el 100 % queda más cerca del centroide A/B; en Manrique Central No.2 (menor riesgo), el 66 %.

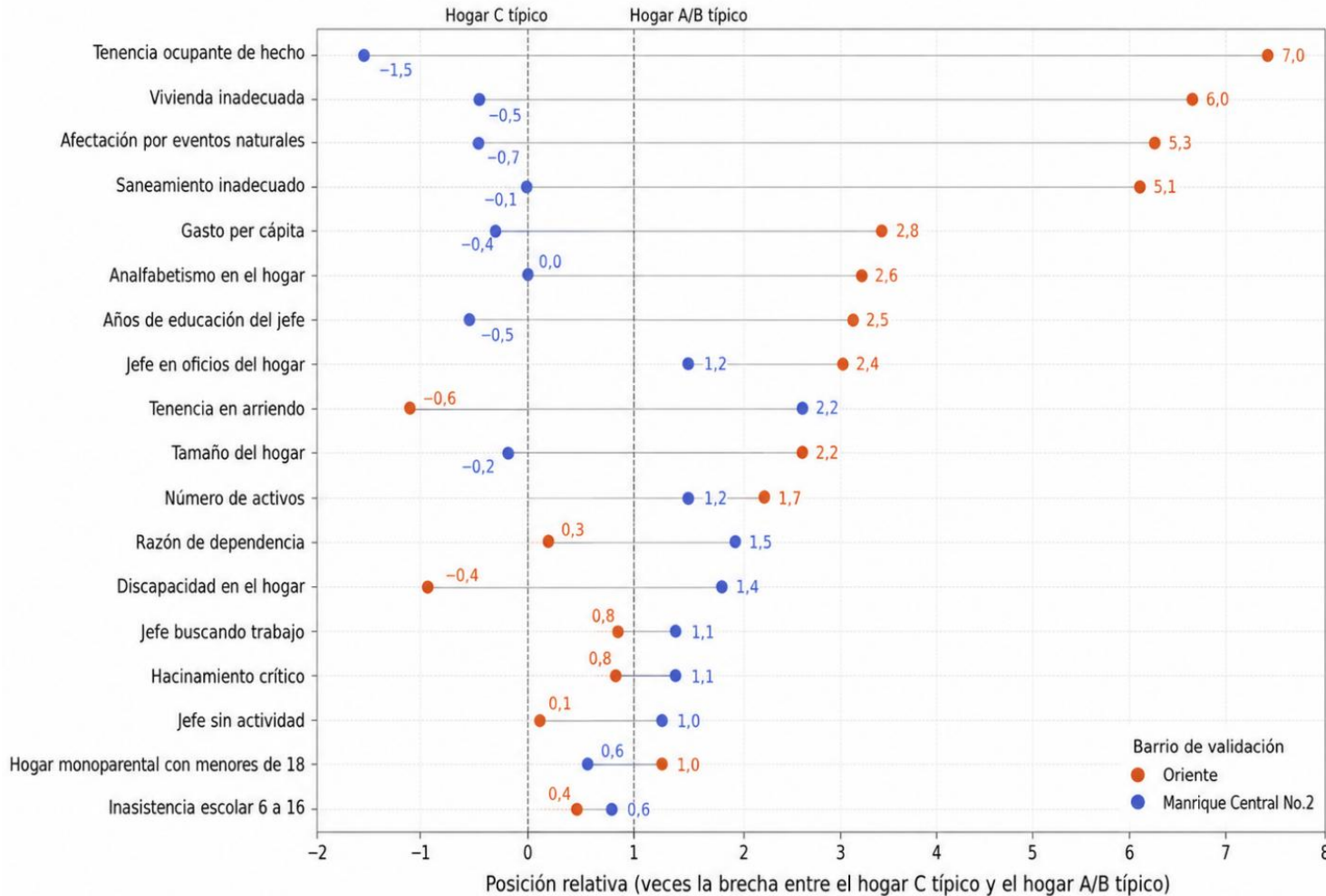
Como referencia, los hogares A/B reales lo hacen en el 70 % de los casos y los C reales solo en el 24 %. Los hogares que el modelo prioriza no solo tienen score alto: su perfil completo se aproxima al de un hogar en pobreza.

Conjunto	n	Score medio	% más cerca del centroide A/B
Hogares A/B reales (referencia)	31.458	—	70 %
Top 100 C — Oriente (312)	100	0,700	100 %
Top 100 C — Manrique Central No.2 (307)	100	0,467	66 %
Hogares C reales (referencia)	62.176	—	24 %



Perfil de los top 100 hogares C por score – Validación de perfil

0 = hogar C típico · 1 = hogar A/B típico · >1 = supera al hogar A/B típico



Perfil comparado de los 100 hogares C

La gráfica compara los 100 hogares C priorizados de dos barrios frente a dos referencias: el hogar C típico, ubicado en 0, y el hogar A/B típico, ubicado en 1. Valores entre 0 y 1 indican que el grupo se acerca al perfil A/B; valores mayores que 1 muestran que, en esa variable, el grupo es incluso más vulnerable que el hogar A/B típico; y valores menores que 0 indican que el grupo está mejor que el hogar C típico en esa dimensión.



Fase 2: Ampliación territorial del MVST, movilidad social e inversión pública

De una lectura focalizada en 3 comunas a una lectura distrital de vulnerabilidad, movilidad social e inversión pública en las 16 comunas y 5 corregimientos de Medellín.

Ampliación del análisis

De 3 comunas piloto a las 16 comunas y 5 corregimientos, con varios cortes del Sisbén IV.

Movilidad social

Traectorias de movilidad ascendente, vulnerabilidad persistente o deterioro relativo entre cortes del Sisbén IV.

Inversión pública territorializada

Tipología territorial :alto riesgo con alta inversión y mejora; alto riesgo con alta inversión, pero sin mejora; alto riesgo con baja inversión; bajo riesgo con alta inversión; y bajo riesgo con baja inversión.

Preguntas orientadoras de la Fase 2

1

¿Qué hogares y territorios presentan movilidad ascendente, vulnerabilidad persistente o deterioro relativo a partir de varios cortes del Sisbén IV?

2

¿Qué territorios concentran hogares C con score alto y, por tanto, requieren seguimiento preventivo?

3

¿Los territorios identificados por el MVST como de mayor riesgo coinciden con los territorios donde se ha concentrado la inversión pública?

4

¿Los territorios que han recibido mayor inversión muestran señales posteriores de mejora, estabilidad o deterioro en los cortes siguientes del Sisbén IV?

Gracias



Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación