



UNIVERSIDAD PONTIFICIA COMILLAS

ICADE — Facultad de Ciencias Económicas y Empresariales

TRABAJO DE FIN DE GRADO

Grado en Business Analytics

**ANÁLISIS DE LA VULNERABILIDAD ECONÓMICA
EN LA POBLACIÓN SENIOR ESPAÑOLA:**

Un enfoque multidimensional basado en la Encuesta de Condiciones de Vida 2025

Autora: Teresa Suquía Vega

Proyecto realizado en colaboración con MAPFRE

Tutor: María Eugenia Fabra Florit

Madrid, 2026

1. Introducción

El envejecimiento de la población española es uno de los principales retos económicos y sociales de las próximas décadas. En 2023, las personas mayores de 65 años superaban ya el 20% de la población española, más de 9,5 millones de personas, y las proyecciones del INE estiman que este porcentaje alcanzará alrededor del 30% antes de 2050 (INE, 2023b). En este trabajo se utiliza el término *hogares seniors* para referirse a aquellos cuya persona de referencia tiene 65 años o más, siguiendo la clasificación de la Encuesta de Condiciones de Vida (INE, 2025). Este cambio no es solo demográfico: cuando la pensión sustituye al salario y el gasto en salud y dependencia crece con la edad, la economía doméstica funciona de otra manera. Y esa otra manera genera presiones que afectan al sistema público de pensiones, a los servicios sanitarios, a la política de vivienda y al mercado asegurador orientado a este segmento (IMSERSO, 2022; OCDE, 2023; ICEA, 2022).

La forma habitual de medir la fragilidad económica en este colectivo es la pobreza monetaria: un hogar está en riesgo cuando su renta equivalente cae por debajo del 60% de la mediana nacional (Eurostat, 2023). Es un criterio útil para comparar entre países, pero tiene un límite claro: no dice nada sobre la capacidad del hogar para aguantar un golpe. Un senior puede estar justo por encima de ese umbral y no tener margen para un gasto sanitario imprevisto, una subida del alquiler o la pérdida del cónyuge. La renta está ahí, el margen no.

La literatura sobre vulnerabilidad parte precisamente de esa distinción. Jalan y Ravallion (1998), Chaudhuri et al. (2002) y Ligon y Schechter (2003) definen la vulnerabilidad no como un estado actual sino como una exposición: la probabilidad de que un shock lleve a un hogar a una pérdida de bienestar que no puede absorber. En la vejez esto cobra especial relevancia porque los ingresos son más rígidos, la capacidad de recomponerse económicamente es menor y los gastos sanitarios y de cuidados tienden a aumentar justo cuando menos se pueden asumir.

Este trabajo analiza la vulnerabilidad económica de los mayores de 65 años en España desde un enfoque multidimensional. La idea no es sustituir la medición de pobreza sino ampliarla: en lugar de mirar solo la renta, se combinan tres señales que recoge la ECV, a saber, ingresos equivalentes por debajo del umbral de pobreza relativa, incapacidad para afrontar gastos imprevistos y dificultad recurrente para llegar a fin de mes. Un hogar senior se clasifica como vulnerable cuando presenta al menos dos de estas tres condiciones

simultáneamente, siguiendo la lógica de identificación multidimensional de Alkire y Foster (2011).

Para responder a esta pregunta se trabaja con los microdatos de la Encuesta de Condiciones de Vida 2025 del INE, con una muestra de 15.974 personas mayores de 65 años. Se combinan tres herramientas: el índice multidimensional de vulnerabilidad, un modelo de regresión logística que estima qué factores aumentan o reducen la probabilidad de ser vulnerable, y un análisis de clustering que agrupa a los seniors en perfiles homogéneos. El resultado no es solo un número de cuántos son vulnerables, sino un mapa de quiénes son y por qué razones distintas llegan a serlo.

La aportación principal es que ese mapa conecta cada perfil con el tipo de instrumento que tiene sentido para él. No es lo mismo un senior vulnerable porque lleva décadas con una pensión insuficiente que uno que era estable hasta que un problema de salud grave le dejó sin margen. El primero necesita garantías de renta; el segundo, protección frente a shocks. Esta diferenciación tiene implicaciones tanto para las políticas públicas como para el mercado asegurador, y la colaboración con MAPFRE permite contrastar los resultados con las necesidades concretas del sector.

El trabajo se organiza de la siguiente manera: el capítulo 2 desarrolla el concepto de vulnerabilidad y lo distingue de la pobreza monetaria; el capítulo 3 revisa los factores que la literatura asocia a la vulnerabilidad en la vejez; el capítulo 4 detalla los datos y la metodología; el capítulo 5 presenta los resultados; y el capítulo 6 cierra con las conclusiones e implicaciones.

2. Vulnerabilidad

2.1. Concepto y delimitación

La vulnerabilidad económica se diferencia de la pobreza en que no mide un estado actual sino una exposición al riesgo: la probabilidad de que un hogar experimente una caída significativa en su bienestar ante un shock que no puede absorber (Chaudhuri et al., 2002; Ligon & Schechter, 2003). Esta distinción entre pobreza observada y vulnerabilidad potencial es el eje conceptual de este trabajo: el objetivo no es identificar quién es pobre hoy, sino quién es frágil ante lo que pueda pasar mañana.

En este estudio, la vulnerabilidad se define operativamente como la situación en que un hogar senior presenta simultáneamente al menos dos de estas tres condiciones: ingresos equivalentes por debajo del umbral de pobreza relativa, incapacidad para afrontar gastos imprevistos, y dificultad recurrente para llegar a fin de mes. Esta definición combina una dimensión objetiva con dos dimensiones subjetivas directamente observables, siguiendo la lógica de identificación dual de Alkire y Foster (2011).

Es importante delimitar qué no mide este concepto. La vulnerabilidad así definida no equivale a exclusión social ni a privación material grave. El análisis se centra en la dimensión de fragilidad financiera, donde los indicadores de la ECV tienen mayor fiabilidad y comparabilidad.

2.2. Por qué la pobreza monetaria no basta

El umbral del 60% de la mediana de ingresos equivalentes (Eurostat, 2023) es útil para comparaciones internacionales, pero tiene cuatro limitaciones sistemáticas como medida exclusiva de la fragilidad de los seniors.

Primera, ignora el patrimonio acumulado: un hogar con ingresos bajos pero con 50.000 euros en depósitos tiene una capacidad de absorción de shocks radicalmente distinta a uno con los mismos ingresos pero sin ahorros. Segunda, su carácter relativo hace que el umbral se desplace con el nivel de vida medio aunque ningún hogar haya cambiado su situación material. En la muestra, el umbral es de 13.096 euros anuales (1.091 euros mensuales). Tercera, no incorpora las necesidades diferenciadas por salud: un hogar con 14.000 euros anuales y enfermedades crónicas tiene necesidades de gasto muy distintas a otro con los mismos ingresos y buena salud. Cuarta, no captura la dinámica: un hogar puede estar sobre

el umbral pero con una trayectoria de recursos decreciente que el indicador estático no detecta.

2.3. Vulnerabilidad estructural y vulnerabilidad sobrevenida

La literatura distingue dos modalidades de vulnerabilidad en la vejez según su origen, y la diferencia es relevante porque condiciona el tipo de intervención que cada una requiere. Schröder-Butterfill y Marianti (2006) proponen un marco que separa la exposición a una amenaza, su materialización y la capacidad de afrontarla, y muestran que esa exposición puede originarse en procesos acumulados a lo largo de la vida o en un evento puntual que la desencadena. La vulnerabilidad estructural es crónica y preexiste a la jubilación: nace de trayectorias laborales con baja capacidad de acumulación que se traducen, décadas después, en pensiones insuficientes para absorber cualquier imprevisto. El hogar nunca llegó a generar un margen real, de modo que la fragilidad acompaña al individuo desde el momento mismo en que se retira del mercado laboral.

La vulnerabilidad sobrevenida, en cambio, afecta a hogares que no eran frágiles hasta que un shock posterior, como el deterioro de la salud, la aparición de dependencia o la pérdida del cónyuge, supera unas reservas que hasta entonces resultaban suficientes. El impacto patrimonial de estos shocks está documentado: Smith (1999) encuentra que la aparición de un problema de salud grave reduce la riqueza del hogar en torno a un 7% de media, una caída que solo se explica en parte por el gasto médico directo y que responde también a la pérdida de ingresos. La trayectoria pudo ser razonable y la pensión adecuada, pero la nueva carga de gastos o la pérdida de ingresos del hogar erosionan el margen presupuestario hasta comprometer su suficiencia.

Esta distinción tiene un paralelismo con la separación entre pobreza crónica y transitoria que proponen Jalan y Ravallion (1998): la primera se atiende mejor con políticas de acumulación de activos, y la segunda con mecanismos de seguro frente a shocks de consumo. Trasladado a este contexto, cada modalidad apunta hacia instrumentos distintos: la vulnerabilidad estructural reclama mecanismos de garantía de rentas, mientras que la sobrevenida es el terreno natural de los productos de protección frente a shocks.

3. Vulnerabilidad en el colectivo senior en España

3.1. Demografía y envejecimiento

España tiene actualmente más de 9,5 millones de personas mayores de 65 años (20,1% de la población), por encima de la media europea del 20,0% y cerca de Italia (23,5%) y Portugal (22,3%) (Eurostat, 2023). Las proyecciones apuntan a que en 2050 ese porcentaje llegará al 31,4% y en 2070 podría alcanzar el 34,6% (INE, 2023b).

El envejecimiento responde a dos tendencias simultáneas: mayor esperanza de vida y caída de la fecundidad. A los 65 años, una mujer española tiene por delante 21,3 años de vida en promedio; un hombre, 18,5 (INE, 2023a). La tasa de fecundidad, en 1,19 en 2022, es una de las más bajas del mundo. El resultado es una pirámide poblacional invertida que estrecha la base de cotizantes sobre la que descansa el sistema de pensiones: el ratio cotizantes/pensionistas ha pasado de 2,6 en 2010 a 2,2 en 2023, y se proyecta en torno a 1,7 para 2040 (Banco de España, 2022).

3.2. La pensión como única fuente de renta

La jubilación supone una ruptura en el modelo de ingresos: el salario, que puede crecer con el tiempo, es sustituido por la pensión pública, que es fija e irreversible al alza. La hipótesis del ciclo vital (Ando y Modigliani, 1963) predice que los hogares suavizan su consumo a lo largo de la vida ahorrando durante la etapa laboral para sostener su nivel de gasto en la jubilación; cuando la pensión no logra esa sustitución, el ajuste recae sobre el consumo presente. Para medir en qué medida la pensión compensa el salario previo se utiliza la tasa de sustitución, que expresa la pensión inicial como porcentaje del último salario. España tiene una tasa media del 71,9% (OCDE, 2023), lo que significa que, de media, un jubilado español cobra una pensión equivalente a casi tres cuartas partes de lo que ganaba trabajando. Ese porcentaje es notablemente superior a la media de los países de la OCDE, que se sitúa en el 51,8%.

Sin embargo, ese dato esconde una heterogeneidad muy alta. La pensión que recibe cada persona depende directamente de su carrera contributiva: cuántos años ha trabajado, con qué salario y en qué régimen. En la muestra analizada, los años totales en trabajo remunerado tienen una media de 33,2 años y una desviación estándar de 14,3, lo que indica que hay personas que apenas han cotizado una década y otras que llevan cuarenta años o

más en el mercado laboral. Esa dispersión se traduce directamente en pensiones muy distintas entre personas del mismo grupo de edad.

La brecha de género es el rasgo más relevante de esa desigualdad. La pensión media de las mujeres fue de 1.065 euros mensuales en 2022, frente a 1.587 euros la de los hombres (Banco de España, 2022), una diferencia del 32,9%. Esa brecha no es un fenómeno de la jubilación en sí, sino el reflejo acumulado de décadas de salarios más bajos, mayor empleo a tiempo parcial y períodos de inactividad dedicados al cuidado familiar, todo lo cual se traduce en carreras contributivas significativamente más cortas y, por tanto, en pensiones más bajas.

3.3. Estructura de gastos en la etapa senior

El gasto sanitario crece de forma no lineal con la edad: los mayores de 75 años generan en promedio un gasto tres veces superior al de la población de 45-54 años (Abellán García et al., 2022). Hay costes que el sistema público no cubre o cubre solo parcialmente. En farmacia, el jubilado paga entre el 10% y el 60% del precio del medicamento según su nivel de renta. La óptica y los audífonos quedan completamente excluidos de la cobertura, igual que las adaptaciones del domicilio para facilitar la movilidad, cuyo coste puede estar entre 3.000 y 20.000 euros.

La dependencia es el riesgo de mayor impacto potencial. Afecta al 5,8% de los seniors de 65-74 años, pero al 43,2% de los mayores de 85 (IMSERSO, 2022), y las proyecciones a largo plazo anticipan que el gasto en cuidados de larga duración en España seguirá creciendo a medida que avance el envejecimiento de la población (Comas-Herrera et al., 2006). El coste de una plaza residencial varía entre 1.500 y 3.000 euros mensuales, con copagos de entre el 25% y el 45% para rentas medias y listas de espera que superaban los 300 días en 2023. En la muestra analizada, el 8,7% presenta limitación funcional severa (PH030 = 1), el indicador adelantado que mejor aproxima el riesgo de dependencia.

Para seniors con ingresos fijos, el gasto en vivienda y suministros supone una fracción creciente del presupuesto. El 14,2% de los hogares con persona de referencia mayor de 65 años no podía mantener su vivienda a temperatura adecuada en 2022, frente al 9,8% de media general (Eurostat, 2023). Esto se explica en parte porque los mayores se concentran en viviendas antiguas, con escaso aislamiento térmico, que resultan más costosas de calentar o refrigerar.

3.4. Diversidad interna del colectivo

Tratar a los mayores de 65 años como un grupo homogéneo es un error metodológico: la literatura gerontológica documenta que la variabilidad entre individuos aumenta con la edad, hasta el punto de que la edad cronológica deja de ser un buen indicador del estado real de una persona mayor (Lowsky et al., 2014). La heterogeneidad interna de este colectivo es comparable a la del conjunto de la población activa y se estructura en torno a cuatro dimensiones.

La primera es la edad. Dentro del colectivo senior conviven personas de 65 años recién jubiladas, con plena autonomía y capacidad de consumo, y personas de 85 o más años con necesidades sanitarias, de cuidado y de adaptación del entorno radicalmente distintas. En la muestra analizada, el 55,1% tiene entre 65 y 74 años, el 31,5% entre 75 y 84, y el 13,4% tiene 85 o más.

La segunda es el género. Las mujeres representan el 54,9% de la muestra, proporción que crece con la edad por su mayor esperanza de vida. Como se ha visto en la sección anterior, sus carreras contributivas son más cortas y sus pensiones significativamente más bajas, lo que las sitúa en una posición estructuralmente más expuesta ante gastos imprevistos.

La tercera es el tipo de hogar. El 14,5% de los mayores en la muestra vive solo. Vivir en solitario elimina la posibilidad de compartir gastos fijos y supone la ausencia de una red de apoyo inmediata ante una enfermedad o una emergencia económica.

La cuarta es el régimen de vivienda. El 81,4% de los mayores en la muestra son propietarios sin hipoteca, lo que representa una fuente de seguridad patrimonial. Sin embargo, el 5,1% son arrendatarios a precio de mercado, expuestos a subidas de alquiler sin posibilidad de absorberlas con un ingreso fijo.

3.5. Factores determinantes de la vulnerabilidad

La revisión de la literatura identifica cinco grupos de factores con poder explicativo consistente, todos incluidos en el modelo de este trabajo.

Nivel de pensión. La pensión pública es el determinante más inmediato de la situación económica del senior. Las carreras cortas o con lagunas de cotización generan pensiones insuficientes para absorber cualquier shock: en la UE, las mujeres reciben de media un 26,1% menos de pensión que los hombres y un 5,3% no percibe ninguna, brechas que se

explican por carreras más cortas e interrumpidas y por el mayor peso del empleo a tiempo parcial (Comisión Europea y Comité de Protección Social, 2024). Se incluye en escala logarítmica (PY100N) para capturar rendimientos decrecientes.

Régimen de vivienda. Los seniors en alquiler presentan tasas de vulnerabilidad consistentemente superiores a los propietarios (Eurostat, 2023). El alquiler opera como una cuña fija entre los ingresos y el consumo disponible, con mayor impacto cuanto más baja es la pensión.

Composición del hogar. Vivir solo implica pérdida de economías de escala y ausencia de red de apoyo informal. La literatura gerontológica documenta que el aislamiento social se asocia con deterioro del bienestar material y subjetivo (Deaton, 2003).

Estado de salud. La relación entre salud e ingresos opera en ambas direcciones: la mala salud genera gastos adicionales y puede limitar la capacidad de trabajo, mientras que los bajos ingresos limitan el acceso a cuidados preventivos (Smith, 1999). Esta endogeneidad impide interpretar los coeficientes del modelo como efectos causales.

Nivel educativo. La educación es el factor estructural con mayor efecto acumulado: determina la calidad del empleo, el nivel salarial, la pensión acumulada y la capacidad de gestión financiera, que Lusardi y Mitchell (2014) vinculan a mejores decisiones de ahorro y planificación de la jubilación. Su impacto se manifiesta décadas antes de la jubilación.

3.6. El mercado de protección financiera

Solo el 0,8% de los mayores de 65 años disponía de cobertura privada de dependencia en 2022 (ICEA, 2022) y apenas el 20,7% de los hogares tenía plan de pensiones privado en 2020 (Banco de España, 2020). Esta baja penetración es coherente con la literatura sobre el mercado de seguros de dependencia, que documenta que la selección adversa derivada de información privada sobre el propio riesgo limita el tamaño de este mercado incluso en países con mayor desarrollo del seguro privado (Finkelstein y McGarry, 2006). Esta baja penetración hace que la segmentación del colectivo —identificar qué tipos de seniors existen y qué necesitan— sea información de alto valor estratégico para el sector asegurador.

3.7. Hipótesis a priori

Antes de ejecutar ningún modelo se formulan seis hipótesis basadas en la literatura revisada en los capítulos 2 y 3. Plantearlas antes de ver los datos tiene dos ventajas: obliga a razonar el porqué antes de ver los resultados, y permite identificar los casos en que la evidencia empírica los contradice —que son, a menudo, los hallazgos más interesantes.

H1. Las mujeres tienen mayor vulnerabilidad por carreras laborales más cortas e interrumpidas que generan pensiones más bajas.

H2. La educación alta protege, a través de empleos más estables, mayores cotizaciones y mejor gestión financiera a lo largo del ciclo vital.

H3. El alquiler de mercado aumenta el riesgo: pagos fijos obligatorios sobre una renta fija, sin el colchón que supone la propiedad.

H4. La mala salud aumenta el riesgo por el gasto sanitario adicional y el riesgo de dependencia.

H5. Vivir solo aumenta el riesgo por la ausencia de economías de escala y de red de apoyo informal.

H6. La pobreza energética es la señal más intensa de vulnerabilidad: no poder calefaccionar la vivienda implica que el gasto en suministros ya consume prácticamente toda la renta disponible.

Estas hipótesis se contrastan con los tests estadísticos (§5.3.5) y con los coeficientes del modelo logístico (§5.5). Los resultados del contraste se recogen en §6.1.

4. Metodología y diseño empírico

4.1. Software

El análisis se realizó en Python 3.12 sobre Google Colab. Librerías principales: pandas 2.1 y numpy 1.26 (manipulación de datos), matplotlib 3.8 y seaborn 0.13 (visualización), scikit-learn 1.4 (Pedregosa et al., 2011; regresión logística, KNN imputer, k-means, métricas) y scipy 1.12 (tests estadísticos, clustering jerárquico).

4.2. Fuente de datos: ECV 2025

El análisis utiliza los microdatos de la Encuesta de Condiciones de Vida (ECV) 2025, elaborada por el INE en el marco armonizado europeo EU-SILC. La encuesta recoge información en dos niveles: el fichero de Hogares (H), con datos sobre vivienda, gastos y privación material, y el fichero de Personas (P), con características demográficas, laborales y de salud de cada miembro. Ambos ficheros se vinculan mediante el identificador HB030. La muestra de personas mayores de 65 años asciende a 16.044 individuos antes de la limpieza y queda en 15.974 tras el tratamiento de valores perdidos descrito en la sección 4.4. La representatividad se garantiza mediante los factores de ponderación transversales (PB040).

4.3. Variables seleccionadas

Se seleccionan 16 variables analíticas, siguiendo dos criterios: relevancia teórica acreditada en la literatura (sección 3.5) y disponibilidad en la ECV 2025 con un porcentaje de missings manejable (máximo 17,8%, en PL200, tratado mediante imputación KNN en la sección 4.4).

Tabla 4.1. Variables analíticas seleccionadas de la ECV 2025

Variable	Fichero	Descripción	Uso en el análisis
HY020	H	Renta disponible total del hogar (€/año)	Dimensión D_INGRESO del índice y predictora
HX240	H	Unidades de consumo (escala OCDE)	Denominador renta equivalente
HS060	H	Capacidad para afrontar gastos imprevistos (1=Sí, 2=No)	Dimensión D_SHOCK del índice
HS120	H	Facilidad para llegar a fin de mes (1–6)	Dimensión D_FINMES del índice
HH021	H	Régimen de tenencia (1–5)	Predictora: alquiler de mercado

Variable	Fichero	Descripción	Uso en el análisis
HH050	H	Puede mantener temperatura adecuada (1=Sí, 2=No)	Predictora: pobreza energética
HH070	H	Gastos de vivienda (€/mes)	Predictora: carga de vivienda
HX060	H	Tipo de hogar (códigos ECV)	Predictora: hogar unipersonal (código 6)
PB140	P	Año de nacimiento	Derivar edad
PB150	P	Sexo (1=Hombre, 2=Mujer)	Predictora: mujer
PB190	P	Estado civil (1–5)	Predictora: viudedad (código 4)
PE041	P	Nivel de estudios (CNED-2014)	Predictora: nivel educativo
PH010	P	Estado de salud (1=muy bueno ... 5=muy malo)	Predictora: mala salud (≥ 4)
PH030	P	Limitación funcional (1=severa, 2=moderada, 3=sin)	Predictora: limitación severa
PY100N	P	Pensión de jubilación neta (€/año)	Predictora: pensión (en log)
PL200	P	Años totales en trabajo remunerado	Predictora: carrera contributiva

Nota: Fichero H = Hogares; Fichero P = Personas. Fuente: INE (2025).

Dos variables requieren precisiones técnicas. HX060 (tipo de hogar): el código 6 corresponde a hogares unipersonales de mayores de 65 años, verificado en los datos donde agrupa a 2.313 seniors. PY100N (pensión): el 30% de la muestra tiene valor cero porque esos seniors no perciben pensión contributiva propia. Se aplica la transformación $\log(1 + PY100N)$ para preservar los ceros y reducir la influencia de valores extremos.

4.4. Tratamiento de valores perdidos

4.4.1. Diagnóstico inicial

Tras convertir todas las variables a formato numérico (`pd.to_numeric` con `errors="coerce"`), el diagnóstico identificó los siguientes porcentajes de missings: PL200 (17,8%), PH030 (4,4%), PH010 (1,0%), PE041 (0,9%), PB190 (0,9%), HH070 (0,3%) y un número residual en HS060, HS120, HH050 y HX060 ($<0,2\%$). La conversión reveló además 49 valores perdidos adicionales en HH070 no detectados inicialmente, confirmando la necesidad de realizar este paso antes del diagnóstico.

Tabla 4.2. Estrategia de imputación por variable

Variable	% Missing	Método	Justificación
PL200	17,8%	KNN (k=5)	Alto % → KNN con auxiliares correlacionadas: PB150, PY100N, HY020, PE041 ($ r >0,10$)
PH030	4,4%	KNN (k=5)	Variables de salud correlacionadas: PH010, edad, PE041, HS120 ($ r >0,15$)
PH010, PE041	~1,0%	Moda por grupo de edad	Pocos missings; patrón generacional claro
PB190	0,9%	Moda global	Pocos missings; sin patrón por cohorte marcado
HH070	0,3%	Mediana por régimen de tenencia	Gasto en vivienda difiere entre propietarios y arrendatarios
HY020 ≤ 0	0,3%	Mediana	Rentas negativas y cero son errores de declaración
HS060/HS120/HH050/HX060	<0,2%	Eliminación de filas	Porcentaje negligible (70 filas, 0,4% de la muestra)

4.4.2. Imputación por KNN

Para PL200 (17,8% de missings) se aplica KNN con $k=5$ porque el alto porcentaje descarta la eliminación de filas. La selección de variables auxiliares se realizó mediante un análisis previo de correlación de Pearson entre PL200 y todas las variables numéricas del dataset: PB150 ($r=-0,491$), PY100N ($r=+0,468$), HY020 ($r=+0,142$) y PE041 ($r=+0,111$). Se aplica un umbral reducido de $|r|>0,10$ —justificado por el alto porcentaje de missings— combinado con criterio conceptual. Para PH030 (4,4%) se aplica el mismo procedimiento, con auxiliares PH010 ($r=-0,645$), edad ($r=-0,279$), PE041 ($r=+0,209$) y HS120 ($r=+0,192$), todas con $|r|>0,15$. En ambos casos, las variables continuas se estandarizan previamente con la escala Z para que las distancias euclídeas no estén dominadas por variables de mayor escala.

4.5. Construcción del índice de vulnerabilidad (VULNERABLE)

4.5.1. Los tres indicadores binarios

El índice se construye a partir de tres indicadores que capturan dimensiones complementarias de la fragilidad financiera.

D_INGRESO = 1 cuando la renta equivalente del hogar (HY020 / HX240, escala OCDE) es inferior al 60% de la mediana ponderada de la muestra senior. La mediana ponderada es de 21.827 euros anuales, lo que sitúa el umbral en 13.096 euros (1.091 euros mensuales).

D_SHOCK = 1 cuando el hogar declara no poder afrontar un gasto imprevisto (HS060 = 2). Captura la fragilidad de liquidez con independencia del nivel de ingresos.

D_FINMES = 1 cuando el hogar declara llegar a fin de mes con dificultad o mucha dificultad ($HS120 \geq 5$). Mide tensión presupuestaria recurrente, no shocks puntuales.

4.5.2. El criterio de acumulación

Se define $VULNERABLE = 1$ si $V = D_INGRESO + D_SHOCK + D_FINMES \geq 2$. El umbral de dos dimensiones simultáneas sigue la lógica de identificación dual de Alkire y Foster (2011) por dos razones. Primera, evita clasificar como vulnerables a hogares con una única dificultad aislada y potencialmente coyuntural. Segunda, hace el indicador más robusto ante el ruido de medición que afecta a cualquier encuesta de autopercepción.

4.6. Modelo de regresión logística

4.6.1. El estimador y su interpretación

La regresión logística modela $P(VULNERABLE = 1) = 1 / (1 + e^{-(\beta_0 + \sum \beta_k X_{ki})})$. A diferencia del modelo de probabilidad lineal, garantiza predicciones acotadas entre 0 y 1. Los coeficientes se interpretan en términos de odds ratios ($OR = \exp(\beta_k)$): un $OR > 1$ indica mayor riesgo relativo de vulnerabilidad, un $OR < 1$ indica efecto protector. Se calculan también los efectos marginales promedio (AME), que miden el cambio en puntos porcentuales en la probabilidad predicha cuando la variable aumenta en una unidad, promediado sobre todas las observaciones.

4.6.2. Variables y estandarización

Se incluyen 13 predictores. Las variables binarias (mujer, mala salud, limitación severa, alquiler, pobreza energética, vivir solo, viudedad) se incluyen directamente. Las tres categorías de educación se codifican con dos dummies, tomando educación baja como referencia. Las cuatro variables continuas (edad, ratio de carga de vivienda, pensión en logaritmo y años trabajados) se estandarizan con la escala Z para comparabilidad de coeficientes. Los errores estándar se calculan por bootstrap con 200 iteraciones.

4.6.3. Diagnóstico de multicolinealidad (VIF)

La multicolinealidad ocurre cuando dos variables del modelo están muy correlacionadas entre sí: el modelo no puede separar bien el efecto de cada una, los errores estándar se inflan y variables relevantes pueden aparecer como no significativas. El Factor de Inflación de la Varianza (VIF) cuantifica este problema para cada variable j : $VIF_j = 1 / (1 - R^2_j)$, donde R^2_j es el coeficiente de determinación de una regresión de la variable j sobre el resto de predictores. $VIF < 5$ indica ausencia de problema; VIF 5-10, multicolinealidad moderada; $VIF > 10$, severa. Los resultados muestran que todas las variables del modelo presentan $VIF < 5$, confirmando la fiabilidad de los coeficientes estimados.

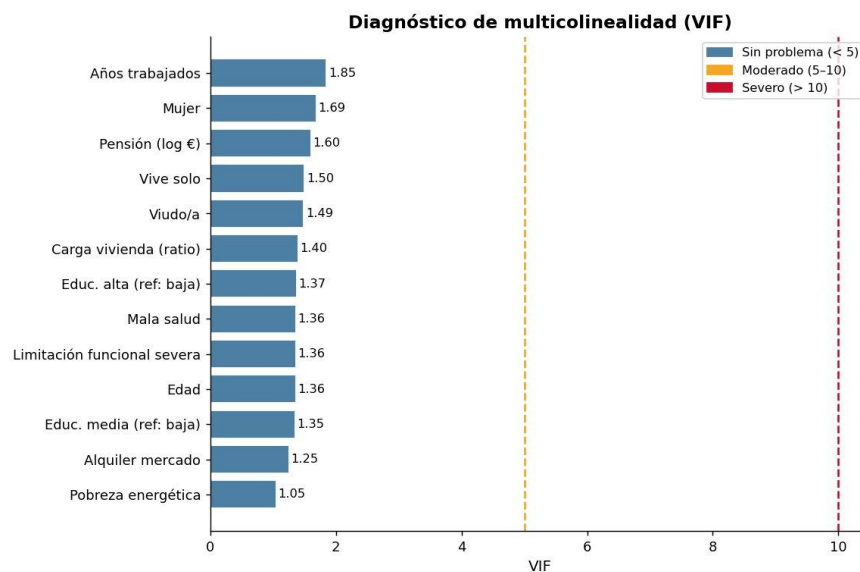


Figura 5. Diagnóstico de multicolinealidad: Factor de Inflación de la Varianza (VIF) por variable. Todas las variables presentan $VIF < 5$.

4.6.4. Métricas de validación

La calidad predictiva se evalúa con el AUC-ROC (0,5 = aleatorio, 0,8 = bueno, 0,9 = excelente), la validación cruzada de 5 pliegues para estimar la capacidad de generalización, y la matriz de confusión para calcular sensibilidad y especificidad con umbral de 0,5.

4.7. Segmentación mediante clustering K-Means

El clustering k-means divide el dataset en k grupos minimizando la suma de distancias euclídeas al cuadrado de cada observación al centroide de su clúster (inercia o WSS). Se utilizan las 13 variables del modelo logístico más los tres indicadores binarios del índice (D_INGRESO, D_SHOCK, D_FINMES). La inclusión de estos últimos es una decisión deliberada: el objetivo del clustering no es replicar de forma independiente la variable VULNERABLE, sino segmentar al colectivo en perfiles internamente homogéneos en

términos de su estructura completa de vulnerabilidad, combinando los factores socioeconómicos con las dimensiones de fragilidad efectivamente observadas. Por construcción, esto introduce una asociación entre los clústeres y la tasa de vulnerabilidad, pero esa asociación es descriptiva, no inferencial: el clustering se utiliza como herramienta de segmentación con fines operativos (diseño de productos e intervenciones), no para estimar el efecto de los predictores sobre la vulnerabilidad, función reservada al modelo logístico. Todas las variables continuas se estandarizan previamente; sin estandarización, variables con mayor varianza dominarían el cálculo de distancias. El número óptimo de clústeres se determina combinando el método del codo y el coeficiente de silueta. Como validación, se realiza un clustering jerárquico con el método de Ward sobre una muestra de 600 observaciones.

5. Resultados

5.1. Estadísticos descriptivos de la muestra

La muestra analizada comprende 15.974 personas mayores de 65 años residentes en España. El 54,9% son mujeres, reflejo de la mayor longevidad femenina. La edad media es de 74,4 años con desviación estándar de 6,8 años.

Tabla 5.1. Estadísticos descriptivos de las variables continuas principales

Variable	Media	Desv. típica	P25	Mediana	P75
Edad (años)	74,4	6,8	68,0	73,0	80,0
Renta equivalente (€/año)	25.127	13.940	15.623	22.858	31.715
Gastos de vivienda (€/mes)	296,8	185,9	193,2	253,4	337,6
Ratio carga vivienda	0,126	0,130	0,060	0,090	0,140
Pensión jubilación neta (€/año)	13.532	12.094	0	11.809	21.983
Años trabajados (PL200)	33,2	14,3	21,2	38,0	44,0

Nota: Estadísticos calculados con factores de ponderación de la ECV 2025. Fuente: elaboración propia a partir de ECV 2025 (INE).

Destaca la dispersión en la pensión de jubilación: el percentil 25 es cero porque aproximadamente el 30% de los seniors no percibe pensión contributiva propia (dependen de la pensión de viudedad del cónyuge o no alcanzaron suficientes años de cotización). Los 126 casos con ratio de carga de vivienda = 1,0 representan hogares donde los gastos de vivienda igualan o superan toda la renta mensual, una situación de estrés financiero extremo.

5.2. Prevalencia de la vulnerabilidad financiera

5.2.1. Uno de cada ocho mayores es vulnerable

El índice identifica a 1.869 seniors en situación de vulnerabilidad sobre 15.974, una tasa del 11,7%. Extrapolando a los 9,5 millones de seniors españoles, equivale a aproximadamente 1,1 millones de personas con fragilidad financiera multidimensional.

Este dato cobra mayor significado cuando se compara con la tasa de pobreza monetaria convencional: el 16,4% de la muestra tiene ingresos por debajo de 13.096 euros anuales. Si

el análisis se hubiera limitado a ese indicador, habríamos identificado 2.612 seniors como en riesgo, pero habríamos pasado por alto a quienes llegan a fin de mes con mucha dificultad o no pueden afrontar un imprevisto sin ser formalmente pobres. El índice multidimensional captura 1.869 personas con fragilidad real en al menos dos dimensiones simultáneas: un grupo más pequeño pero con señales más sólidas de vulnerabilidad genuina.

5.2.2. Las tres dimensiones: dónde está la fragilidad

La tensión presupuestaria mensual (D_FINMES) es la dimensión más prevalente: el 32,8% de los seniors declara llegar a fin de mes con dificultad o mucha dificultad. Le sigue la ausencia de colchón financiero (D_SHOCK), con el 26,6%. La dimensión monetaria (D_INGRESO) es la tercera, con el 16,4%. A estas tres dimensiones que componen el índice se añade, como predictor relacionado, la pobreza energética (HH050), que afecta al 12,1% de la muestra: seniors cuyo hogar no puede mantenerse a temperatura adecuada en invierno. El dato más revelador: el 29,1% de la muestra declara tensión mensual sin estar por debajo del umbral de renta, son seniors que "no son pobres" pero que cada mes sienten que el dinero no llega. Esta desconexión es la principal justificación empírica para el enfoque multidimensional.

5.2.3. La distribución del índice V

El 36,3% de los seniors no tiene ninguna fragilidad observable (V=0). El 52,0% tiene exactamente una (V=1): son frágiles ante un shock adicional, pero no vulnerables en sentido multidimensional. Los clasificados como VULNERABLE=1 son quienes tienen dos dimensiones activas (V=2, 11,4%). Y los 53 seniors con las tres simultáneas (V=3, 0,3%) representan el núcleo duro: ingresos bajos, sin colchón y con tensión mensual persistente al mismo tiempo.

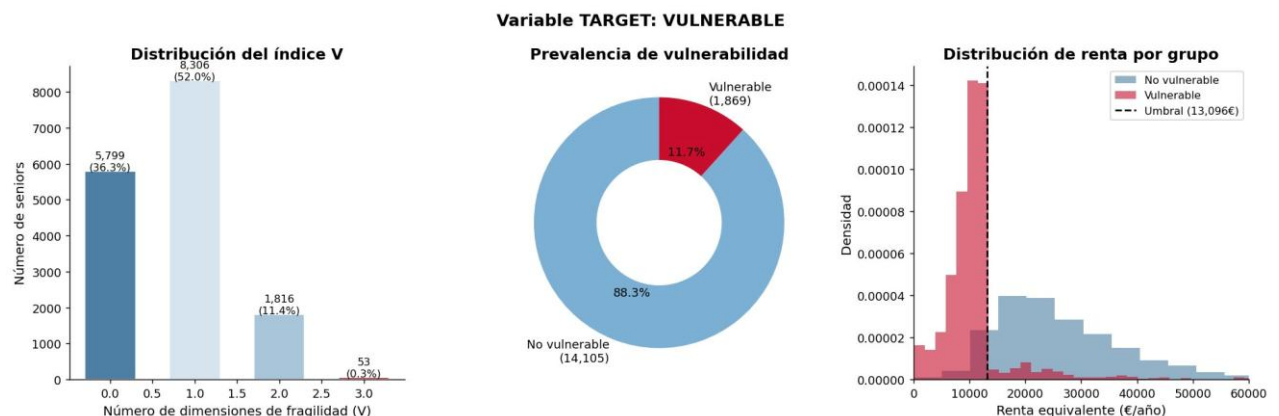


Figura 1. Distribución del índice V, prevalencia de vulnerabilidad (11,7%) y distribución de renta equivalente por grupo.

5.3. Análisis descriptivo por subgrupos

El análisis de la tasa de vulnerabilidad por subgrupos revela dónde se concentra la fragilidad. Las diferencias son, en algunos casos, muy pronunciadas.

Tabla 5.2. Tasa de vulnerabilidad financiera por subgrupos demográficos y del hogar

Subgrupo	N	Vulnerables	Tasa (%)
Grupo de edad			
65-74 años	8.802	944	10,7
75-84 años	5.024	634	12,6
85 o más años	2.148	291	13,5
Sexo			
Hombre	7.210	715	9,9
Mujer	8.764	1.154	13,2
Nivel educativo			
Bajo (sin estudios/primaria)	6.582	1.068	16,2
Medio (secundaria/bachiller)	6.083	622	10,2
Alto (FP sup./universitario)	3.309	179	5,4
Régimen de tenencia			
Propiedad sin hipoteca	13.008	1.331	10,2
Propiedad con hipoteca	1.260	114	9,0
Alquiler mercado	808	202	25,0
Alquiler precio reducido	328	85	25,9
Cesión gratuita	570	137	24,0
Tipo de hogar			
Unipersonal (65+)	2.313	533	23,0
Pareja con otros	11.626	988	8,5
Pareja sin dependientes	1.041	172	16,5
Hogar complejo	994	176	17,7
Estado de salud			

Subgrupo	N	Vulnerables	Tasa (%)
Muy buena / Buena	7.692	698	9,1
Regular	5.877	788	13,4
Mala / Muy mala	2.405	383	15,9

Fuente: elaboración propia a partir de ECV 2025 (INE).

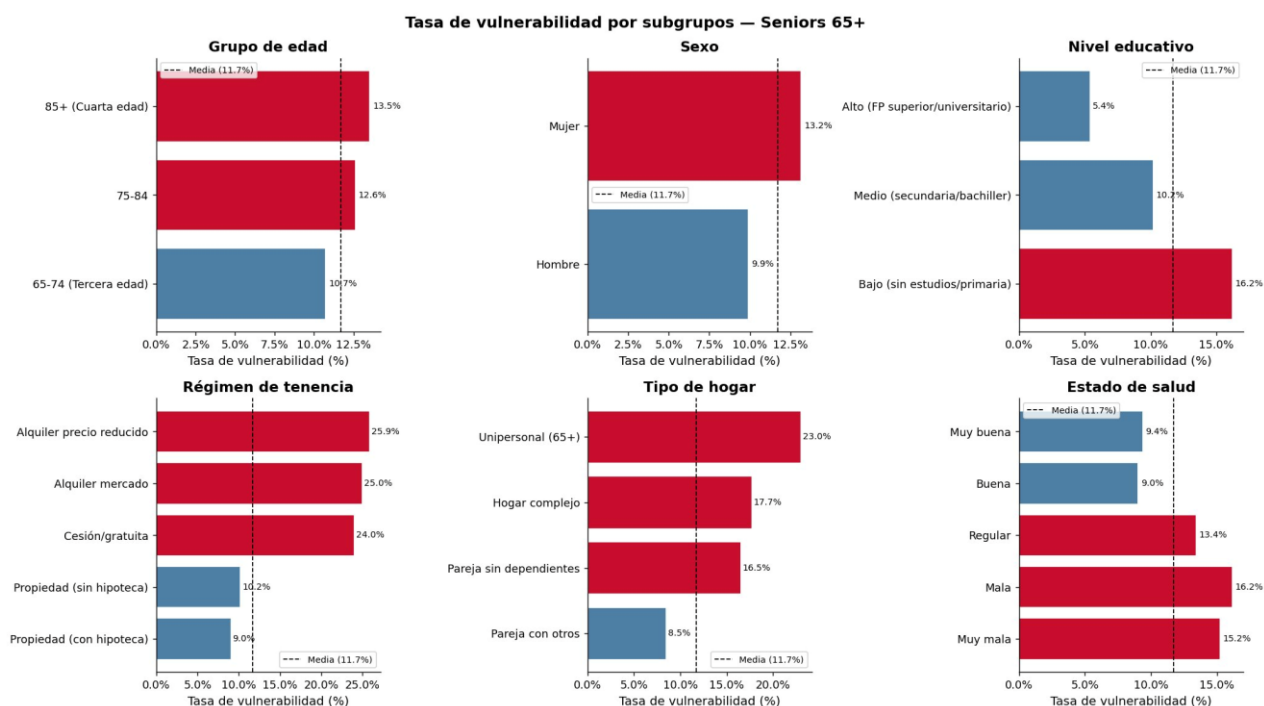


Figura 2. Tasa de vulnerabilidad financiera por subgrupos demográficos y del hogar.

5.3.1. La educación: la inversión que nunca caduca

El gradiente educativo es el más pronunciado de todos los subgrupos: la tasa pasa del 16,2% en los seniors con educación baja al 5,4% en los de educación alta. Un senior con estudios universitarios tiene un riesgo tres veces menor que uno sin estudios o con solo primaria. La explicación es acumulativa: la educación determina la estabilidad y remuneración del empleo a lo largo de la vida laboral, lo que se traduce en pensiones más altas que actúan como escudo frente a las tres dimensiones de fragilidad simultáneamente.

5.3.2. La vivienda en alquiler: pagar cada mes con ingresos fijos

Los seniors en alquiler de mercado presentan una tasa del 25,0%, más del doble que los propietarios sin hipoteca (10,2%). El mecanismo es simple: un jubilado con 1.000 euros de pensión que destina 600 al alquiler tiene 400 para todo lo demás, sin ninguna posibilidad de aumentar los ingresos si el casero sube el precio. Los arrendatarios de precio reducido

(25,9%) y los hogares en cesión gratuita (24,0%) presentan tasas similares, lo que sugiere que estos regímenes se asocian a condiciones socioeconómicas especialmente desfavorables.

5.3.3. Vivir solo: sin red de seguridad ni economías de escala

Los hogares unipersonales presentan una tasa del 23,0%, casi tres veces la de los hogares de pareja con otros miembros del hogar (8,5%). Operan dos mecanismos simultáneos: la ausencia de economías de escala (el alquiler, el gas o el transporte no se comparten) y la falta de red de apoyo informal (un senior solo que rompe la lavadora o necesita ir al médico no tiene a nadie que absorba ese shock). El 14,5% de los seniors vive solo; su tasa de vulnerabilidad triplica la de quienes conviven en pareja con otros miembros del hogar.

5.3.4. Salud y género

El estado de salud muestra el gradiente esperado: del 9,1% en los de salud muy buena o buena al 15,9% en los de salud mala o muy mala. Cada escalón de deterioro añade costes de farmacia, transporte sanitario y adaptación del hogar que, en un presupuesto fijo, reducen el margen. En cuanto al género, las mujeres presentan una tasa del 13,2% frente al 9,9% de los hombres. Esta diferencia de 3,3 puntos parece modesta en el descriptivo, pero esconde disparidades grandes cuando se combina con otros factores: la mujer de 80 años que vive sola (Clúster 0) tiene una tasa del 17% y una probabilidad predicha individual de hasta el 58%.

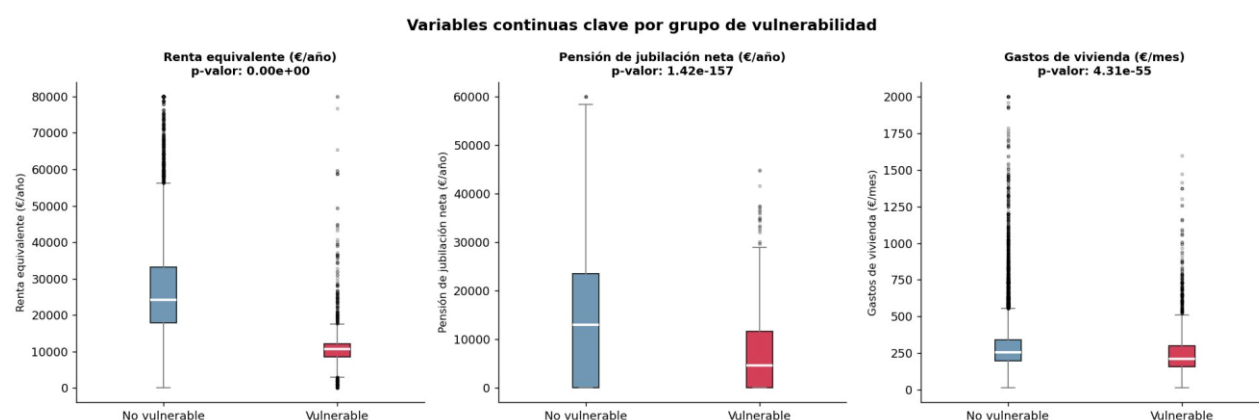


Figura 3. Distribución de renta equivalente, pensión y gastos de vivienda por grupo de vulnerabilidad (test de Mann-Whitney, $p < 0,001$ en los tres casos).

5.3.5. Tests estadísticos de asociación

Para confirmar que las asociaciones observadas en los gráficos y tablas son reales y no fruto del azar, se aplican tests estadísticos formales. Para variables binarias y categóricas se usa

el test χ^2 de Pearson, que comprueba si la distribución de vulnerables difiere significativamente entre categorías. La fuerza de la asociación se mide con la V de Cramér (de 0 a 1, independiente del tamaño muestral). Para variables continuas se usa el test de Mann-Whitney U, alternativa no paramétrica al t-test que no asume normalidad, una condición que raramente cumplen variables económicas como la pensión o la renta. La magnitud del efecto se mide con la r biserial. El nivel de significación es $\alpha = 0,05$.

Todos los predictores presentan asociaciones estadísticamente significativas con la vulnerabilidad ($p < 0,001$). Los efectos más fuertes son los de renta equivalente (r biserial = 0,48) y pensión de jubilación (r biserial = 0,41). Entre las variables binarias, la pobreza energética ($V = 0,34$) y el alquiler de mercado ($V = 0,22$) son las más discriminantes. La edad muestra asociación positiva y significativa en el bivariante (r biserial = 0,07), pero como se verá en el modelo, ese efecto desaparece cuando se controla por el resto de variables.

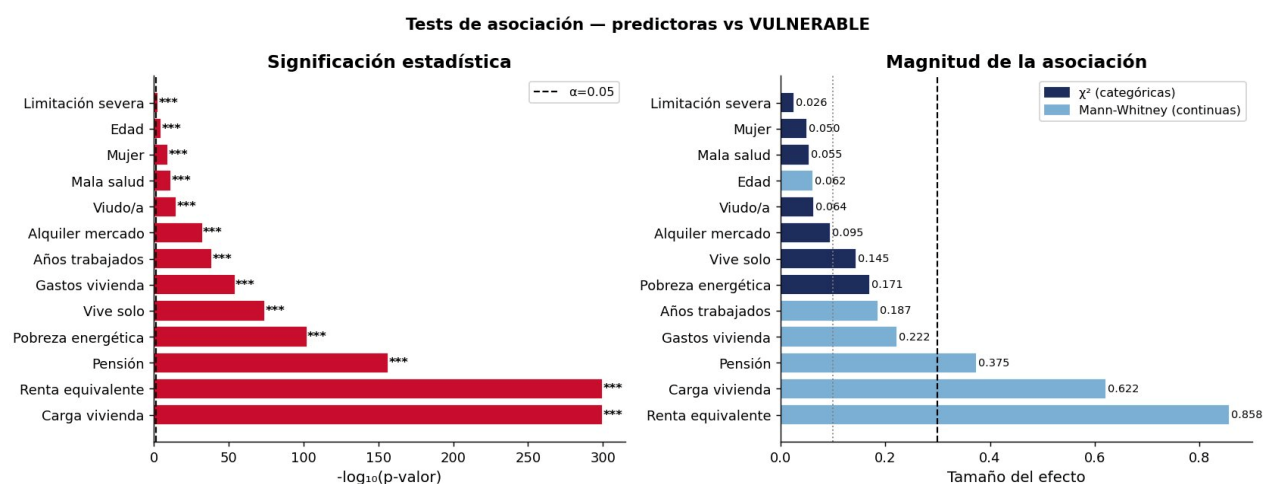


Figura 4. Significación estadística (izquierda) y tamaño del efecto (derecha) de las asociaciones entre cada predictor y VULNERABLE.

5.4. Matriz de correlaciones

La matriz de correlaciones permite leer de un vistazo las asociaciones bivariantes más importantes antes de entrar en el análisis multivariante. La correlación más alta es la de D_INGRESO con VULNERABLE ($r=0,73$): la dimensión monetaria es la que más contribuye al índice. D_SHOCK también correlaciona positivamente ($r=0,45$), mientras que D_FINMES presenta una correlación prácticamente nula con VULNERABLE pese a formar parte del índice. Esto se explica por la propia construcción del indicador: D_FINMES es la dimensión con mayor prevalencia (32,8%) y un porcentaje elevado de

hogares la presenta de forma aislada, sin acumular D_INGRESO ni D_SHOCK, por lo que no llegan a clasificarse como vulnerables ($V \geq 2$). En la práctica, la tensión mensual funciona como una señal de baja especificidad: aparece tanto en hogares verdaderamente frágiles como en hogares que llegan ajustados a fin de mes sin estar en situación de fragilidad multidimensional. Esta característica refuerza, precisamente, la decisión metodológica de exigir al menos dos dimensiones simultáneas para identificar la vulnerabilidad.

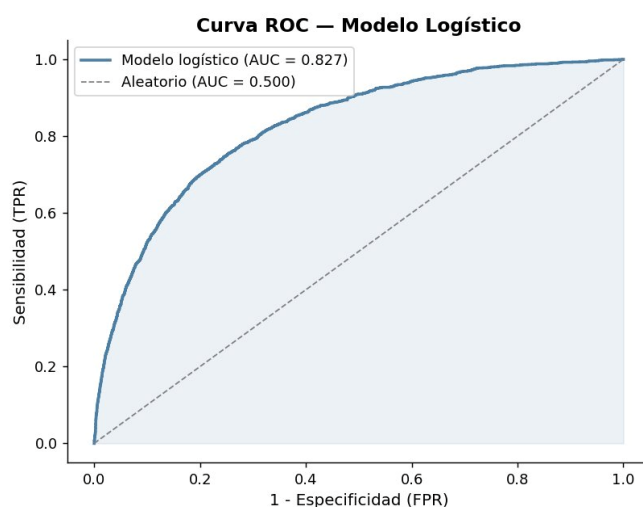
La correlación más visible en el otro extremo es la de "Años trabajados" con "Mujer" ($r = -0,55$): las mujeres de la muestra tienen en promedio 20 años menos de carrera laboral que los hombres. Este número captura en una sola cifra la brecha de género del mercado laboral español de las décadas pasadas.

5.5. Determinantes de la vulnerabilidad: modelo logístico

5.5.1. Capacidad predictiva del modelo

El modelo logístico alcanza un AUC de 0,827: si tomamos al azar un senior vulnerable y uno no vulnerable, el modelo asigna al vulnerable una probabilidad predicha mayor en el 82,7% de los casos. La validación cruzada de 5 pliegues confirma la robustez: AUC promedio $0,825 \pm 0,018$, prácticamente idéntico. El modelo generaliza bien.

Un matiz sobre la matriz de confusión: con umbral de 0,5, la sensibilidad es solo del 15,7%. Esto no indica mal modelo sino la consecuencia natural del desbalance de clases (11,7% de vulnerables): el modelo solo clasifica como vulnerable quien tiene señales muy claras. La especificidad (98,7%) confirma que casi nunca etiqueta erróneamente a un no vulnerable. Para aplicaciones de detección activa, sería aconsejable bajar el umbral a 0,2-0,3.



Matriz de confusión (umbral = 0.5)

True label	No vulnerable	Vulnerable
	13928	177
Vulnerable	1575	294
	No vulnerable	Vulnerable
Predicted label		

Figura 6. Curva ROC ($AUC = 0,827$) y matriz de confusión con umbral de 0,5.

5.5.2. Factores de riesgo y protectores

Tabla 5.3. Resultados del modelo logístico: odds ratios, intervalos de confianza y efectos marginales

Variable	OR	IC 95%	AME (pp)	Significación
Pobreza energética	2,395	[2,102; 2,728]	+7,21	*** ($p<0,001$)
Carga vivienda (ratio)	2,289	[2,078; 2,521]	+6,84	*** ($p<0,001$)
Vive solo	1,899	[1,595; 2,261]	+5,30	*** ($p<0,001$)
Mala salud	1,215	[1,045; 1,413]	+1,61	* ($p<0,05$)
Edad (estánd.)	0,992	[0,930; 1,057]	-0,07	n.s. ($p=0,80$)
Limitación severa	0,894	[0,727; 1,101]	-0,92	n.s. ($p=0,29$)
Años trabajados (estánd.)	0,860	[0,799; 0,925]	-1,25	*** ($p<0,001$)
Viudo/a	0,776	[0,654; 0,922]	-2,09	** ($p<0,01$)
Pensión (log €, estándar.)	0,761	[0,713; 0,813]	-2,25	*** ($p<0,001$)
Mujer	0,573	[0,489; 0,670]	-4,60	*** ($p<0,001$)
Educ. media (ref: baja)	0,557	[0,491; 0,631]	-4,84	*** ($p<0,001$)
Alquiler mercado	0,444	[0,320; 0,615]	-6,71	*** ($p<0,001$)
Educ. alta (ref: baja)	0,272	[0,221; 0,335]	-10,76	*** ($p<0,001$)

Nota: *** $p<0,001$; ** $p<0,01$; * $p<0,05$; n.s. no significativo. OR = odds ratio; AME = efecto marginal promedio en puntos porcentuales. Variables continuas estandarizadas. Errores estándar por bootstrap (200 iteraciones). Fuente: elaboración propia a partir de ECV 2025 (INE).

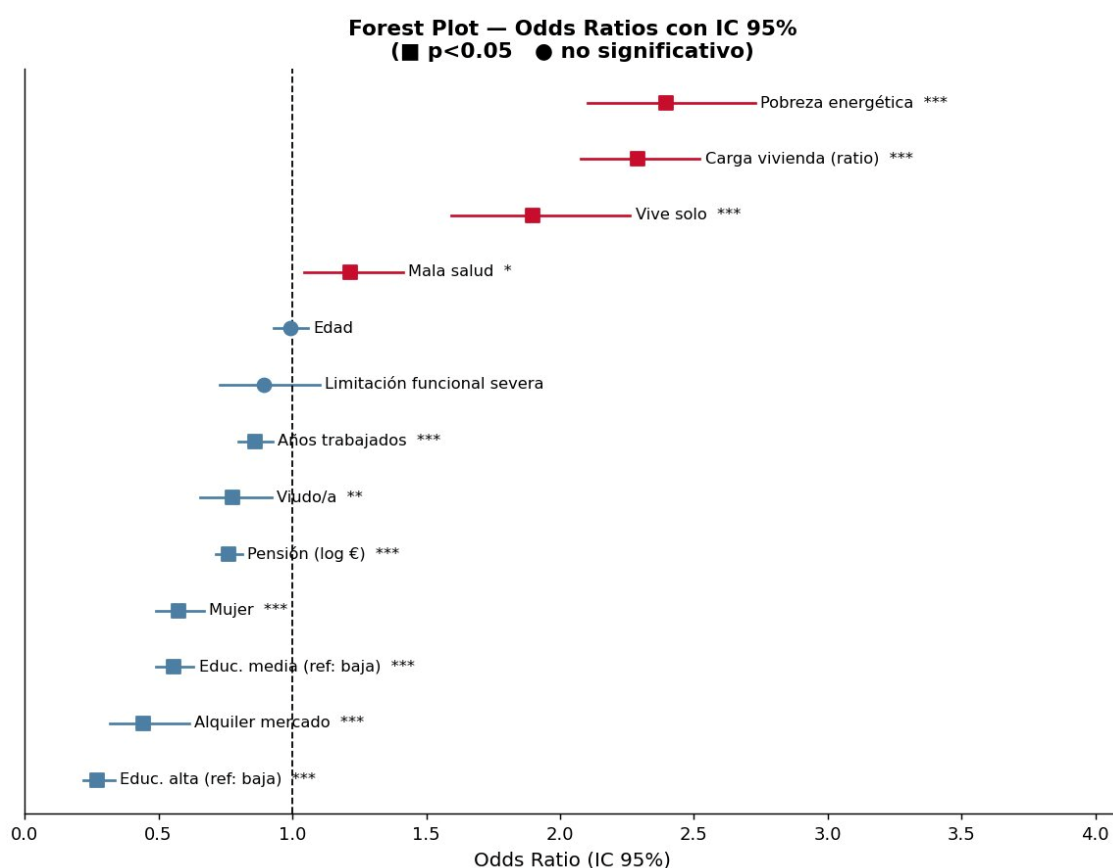


Figura 7. Forest plot de odds ratios con intervalos de confianza al 95%. En rojo, factores de riesgo ($OR > 1$); en azul, factores protectores ($OR < 1$). Los círculos indican no significativo; los cuadrados, $p < 0,05$.

La pobreza energética es el predictor de riesgo más potente ($OR=2,395$, $AME=+7,21$ pp). Un senior que en invierno no puede mantener su vivienda a temperatura adecuada tiene más del doble de probabilidad de ser vulnerable. No poder pagar la calefacción es a la vez la consecuencia de tener poco dinero y la señal de que ese dinero no alcanza para nada más.

La carga de vivienda ($OR=2,289$, $AME=+6,84$ pp) y vivir solo ($OR=1,899$, $AME=+5,30$ pp) son los siguientes factores de riesgo. El efecto de vivir solo confirma que la vulnerabilidad no es solo de renta: los ingresos se controlan en el modelo, pero un senior solo que rompe la lavadora no tiene a nadie que absorba ese shock.

La mala salud tiene $OR=1,215$ y $AME=+1,61$ pp. El efecto es menor de lo esperado porque el índice VULNERABLE ya captura parte del impacto de la mala salud a través de D_FINMES y D_SHOCK: los enfermos ya aparecen como más frágiles en esas dimensiones, de modo que el efecto neto una vez controladas es más acotado.

La edad no es estadísticamente significativa ($p=0,80$). No porque no importe, sino porque sus consecuencias, pensión baja, mala salud, viudedad, vivir solo, ya están controladas por otras variables del modelo. Una vez que sabemos el nivel de pensión, si vive solo y cuántos años trabajó, saber que tiene 80 en lugar de 70 no añade información. Esto tiene una implicación práctica directa: diseñar productos de protección únicamente por edad es ineficiente.

El factor protector más potente es la educación alta ($OR=0,272$, $AME=-10,76$ pp): reduce en más del 70% las probabilidades de vulnerabilidad respecto a educación baja. Para contextualizar: pasar de educación baja a alta tiene casi el mismo efecto (en magnitud opuesta) que acumular simultáneamente pobreza energética, alta carga de vivienda y hogar unipersonal.

5.5.3. Los resultados contraintuitivos: género y alquiler

Ser mujer ($OR=0,573$) y tener alquiler de mercado ($OR=0,444$) aparecen como factores protectores en el modelo, aunque el análisis descriptivo muestra que mujeres y arrendatarios tienen tasas de vulnerabilidad más altas que la media. La aparente contradicción se resuelve con el control estadístico.

El modelo no compara mujeres con hombres en general: compara una mujer y un hombre con exactamente la misma pensión, la misma carrera laboral y el mismo tipo de hogar. Cuando se equiparan esas condiciones, las mujeres gestionan mejor el presupuesto. Son más vulnerables en el descriptivo porque tienen pensiones más bajas y viven más frecuentemente solas —efectos ya capturados por PY100N y HX060 en el modelo—, no porque ser mujer per se incremente la fragilidad.

La misma lógica aplica al alquiler: una vez que el modelo controla por pensión y ratio de carga de vivienda, el efecto neto del alquiler como categoría es negativo, porque los seniors en alquiler de mercado tienden a concentrarse en zonas urbanas con pensiones más altas. Pero, y esto es crucial, cuando el alquiler se combina con pensión baja y ausencia de colchón, el resultado es el Clúster 3, con la mayor tasa de vulnerabilidad del colectivo (31,8%). El modelo captura efectos marginales individuales; el clustering muestra las combinaciones realmente peligrosas.

5.6. Perfiles hipotéticos de riesgo

Para hacer más tangibles las predicciones del modelo, se construyeron seis perfiles hipotéticos representativos de distintas combinaciones frecuentes entre los seniors españoles.

Tabla 5.4. Probabilidades predichas de vulnerabilidad para perfiles hipotéticos

Perfil	Características principales	P(VULNERABLE)
1 — Máximo riesgo	Mujer, 80+, sola, alquiler, mala salud, 15 años trabajados	58,3%
2 — Riesgo alto	Mujer, 70-75, sola, alquiler, salud regular, 20 años trabajados	25,9%
3 — En el límite	Hombre, 75, viudo, propiedad, mala salud, 38 años trabajados	13,3%
4 — Riesgo bajo	Hombre, 70, pareja, propiedad, buena salud, 40 años trabajados	8,1%
5 — Riesgo muy bajo	Mujer, 68, pareja, propiedad, buena salud, 25 años trabajados	4,8%
6 — Mínimo riesgo	Hombre, 65, pareja, propiedad, muy buena salud, 45 años, pensión alta	2,8%

Nota: Probabilidades calculadas aplicando el modelo logístico estimado. Fuente: elaboración propia a partir de ECV 2025 (INE).

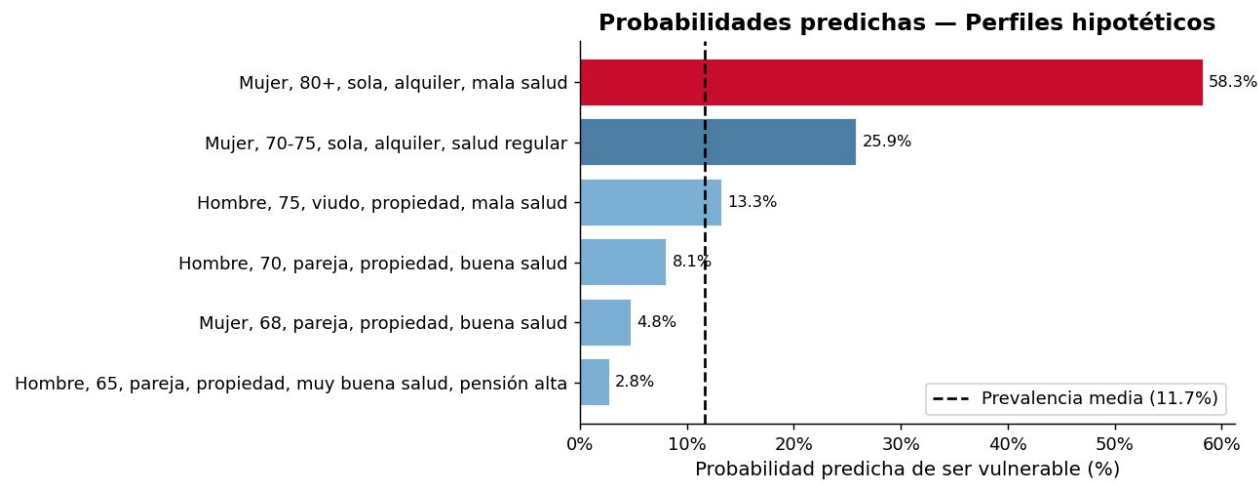


Figura 8. Probabilidades predichas de ser vulnerable para seis perfiles hipotéticos representativos del colectivo senior español.

El Perfil 1 —mujer de más de 80 años, sola, en alquiler, con mala salud y solo 15 años de trabajo— tiene una probabilidad predicha del 58,3%. Este es el retrato estadístico de la vulnerabilidad más extrema: una mujer que dedicó su vida activa al cuidado familiar y que

ahora se enfrenta sola a una pensión mínima, una enfermedad crónica y un alquiler que consume más de la mitad de sus ingresos. No es un perfil excepcional: como se verá en la segmentación de la sección 5.7, este combinado de características agrupa a varios miles de personas dentro del colectivo senior español.

La distancia entre el Perfil 1 (58,3%) y el Perfil 6 (2,8%), con 55,5 puntos porcentuales, ilustra la enorme heterogeneidad del riesgo dentro del colectivo. No existe "el senior": existe un continuo de situaciones desde la fragilidad extrema hasta la estabilidad sólida.

El Perfil 3 (hombre viudo de 75 años con propiedad y mala salud, 13,3%) es especialmente ilustrativo porque se sitúa justo en la media del colectivo (11,7%). La mala salud deteriora la situación, pero la propiedad de la vivienda y la carrera laboral larga la compensan. Ningún factor actúa en el vacío.

5.7. Segmentación del colectivo senior: clustering

5.7.1. Selección del número de clústeres

El método del codo muestra una desaceleración clara en la reducción de inercia a partir de $k=4$: entre $k=3$ y $k=4$ la mejora es sustancial; a partir de $k=5$ es marginal. El coeficiente de silueta confirma $k=4$ como óptimo (0,173), con una caída abrupta a $k=5$ (0,151). El dendrograma jerárquico valida visualmente la estructura de cuatro ramas principales bien diferenciadas.

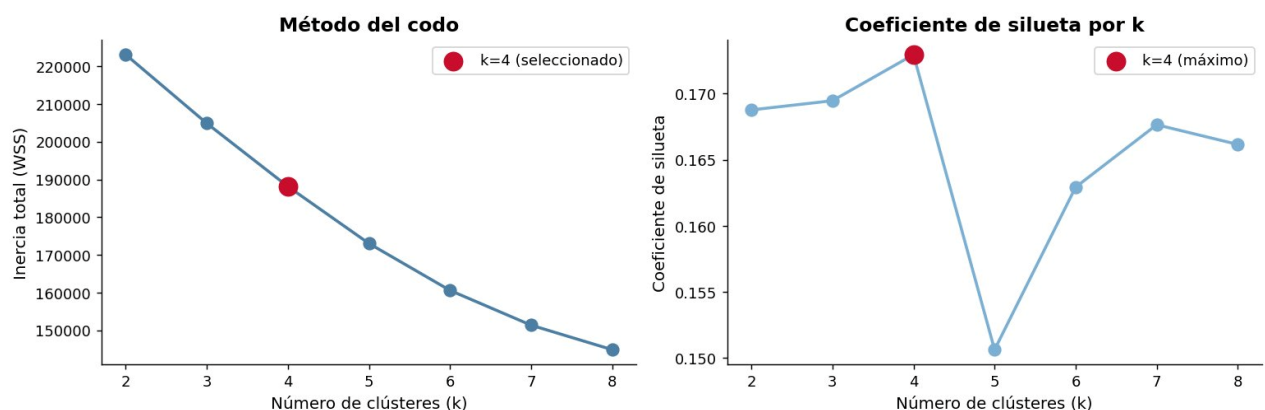


Figura 9. Selección del número óptimo de clústeres: método del codo (izquierda) y coeficiente de silueta (derecha). Ambos criterios convergen en $k=4$.

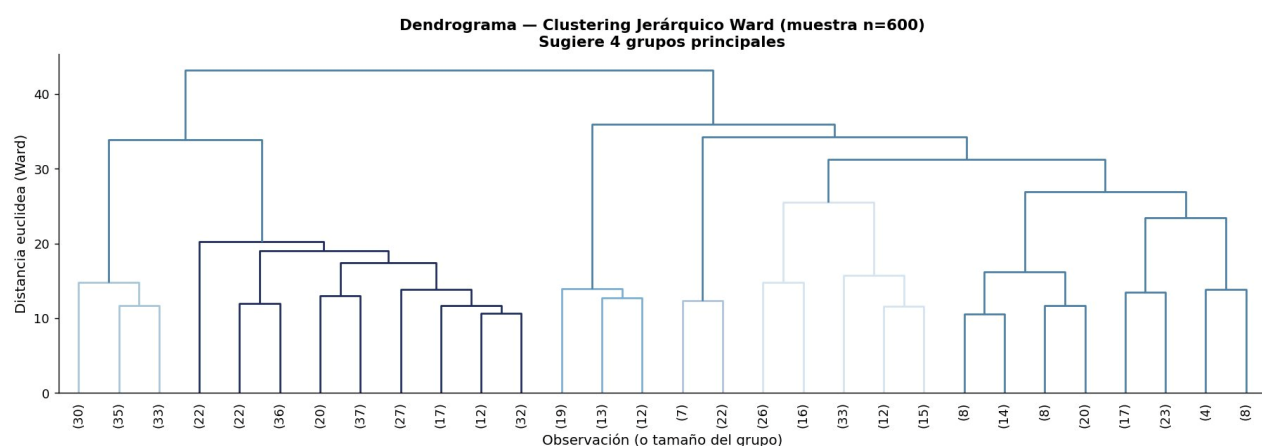


Figura 10. Dendrograma del clustering jerárquico Ward (muestra n=600). Las cuatro ramas principales confirman la estructura de $k=4$.

5.7.2. Cuatro perfiles, cuatro historias

Tabla 5.5. Perfiles de los cuatro clústeres identificados mediante k -means ($k=4$)

Indicador	Clúster 0	Clúster 1	Clúster 2	Clúster 3
N (personas)	5.126	1.332	8.591	925
% de la muestra	32,1%	8,3%	53,8%	5,8%
% VULNERABLE	17,0%	13,7%	6,0%	31,8%
Edad media (años)	75,6	79,2	73,1	73,4
% Mujeres	97,8%	63,7%	27,7%	56,9%
% Vive solo	33,2%	16,9%	2,2%	21,2%
% Viudo/a	42,5%	39,3%	7,2%	21,2%
% Alquiler mercado	0,0%	1,5%	0,0%	85,2%
% Mala salud	13,5%	73,3%	7,0%	14,7%
% Limitación severa	0,0%	99,6%	0,0%	6,2%
% Pobreza energética	15,0%	20,6%	7,8%	24,0%
Años trabajados (media)	20,6	30,2	41,4	32,5
% Ingreso bajo	24,8%	19,9%	8,5%	37,8%
% Sin colchón	32,6%	37,9%	18,3%	54,1%
% Tensión mensual	26,2%	20,3%	40,1%	19,4%

Nota: k -means con $k=4$, $n_init=20$, $random_state=42$. Silueta=0,173. Fuente: elaboración propia a partir de ECV 2025 (INE).

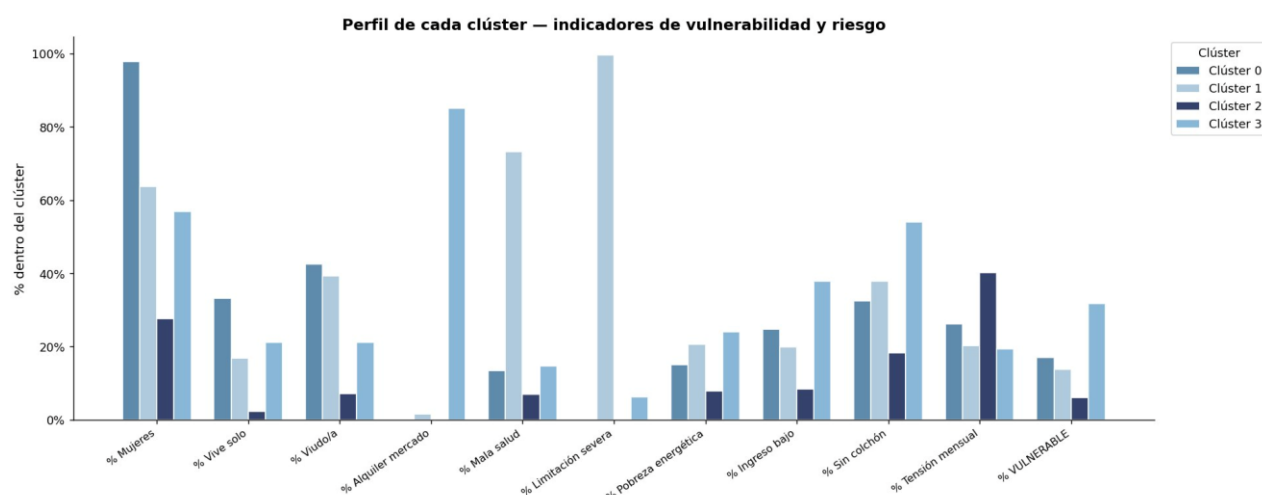


Figura 11. Perfil comparativo de los cuatro clústeres en indicadores de vulnerabilidad y riesgo.

El Clúster 0 —las mujeres mayores solas— agrupa a 5.126 personas (32,1% de la muestra). El 97,8% son mujeres, con una media de solo 20,6 años de trabajo remunerado: la mujer típica trabajó en el mercado formal hasta los 30-35 años y dedicó el resto de su vida activa a tareas de cuidado no remuneradas. El resultado, décadas después, es una pensión baja o inexistente. Lo llamativo es que casi nadie en este clúster vive en alquiler de mercado (0%): casi todas son propietarias. Su vulnerabilidad no viene del alquiler sino de pensiones insuficientes y de la ausencia de red de apoyo. Con 800-900 euros de pensión media, los gastos ordinarios ya consumen casi todo.

El Clúster 1 —dependencia funcional— agrupa a 1.332 personas (8,3%). El 99,6% tiene limitación funcional severa y el 73,3% declara mala salud. Es la imagen del envejecimiento avanzado en su expresión más difícil: personas de 79,2 años de media que no pueden realizar actividades básicas sin ayuda. Este clúster ilustra la vulnerabilidad sobrevenida: no se llega a él necesariamente por haber tenido bajos ingresos (los años trabajados medios, 30,2, son superiores a los del Clúster 0), sino por el deterioro de una situación que en algún momento fue estable. El 37,9% no puede afrontar imprevistos, y con ese perfil de salud los imprevistos son frecuentes: una caída, una infección, adaptar el baño, contratar un cuidador.

El Clúster 2 —seniors protegidos— es el más grande con 8.591 personas (53,8%). Con tasa de vulnerabilidad del 6,0%, representa al senior español tipo: hombre (72,3%) con carrera laboral larga (41,4 años de media), propietario, sin limitaciones funcionales. Sus factores de riesgo potenciales, deterioro de salud, pérdida del cónyuge, serían manejables desde su posición económica actual.

El Clúster 3 —vulnerabilidad estructural por alquiler— es el más pequeño (925 personas, 5,8%) pero el de mayor tasa de vulnerabilidad (31,8%), casi tres veces la media. El 85,2% vive en alquiler de mercado. La situación concreta: un jubilado en una gran ciudad española, con 900-1.000 euros de pensión, que paga 650-700 de alquiler. Después del alquiler quedan 250-350 euros para todo lo demás. Es matemáticamente imposible ahorrar, y cualquier imprevisto supera de inmediato los recursos disponibles. El 54,1% no puede afrontar imprevistos, la proporción más alta de los cuatro grupos, y el 37,8% tiene ingresos por debajo del umbral de pobreza.

5.7.3. Cuatro grupos, cuatro necesidades distintas

Los cuatro clústeres no son abstracciones estadísticas: son retratos de perfiles reales con necesidades de protección financiera específicas que ningún producto genérico puede atender eficientemente. El análisis proporciona un mapa de segmentos sobre el que diseñar respuestas diferenciadas.

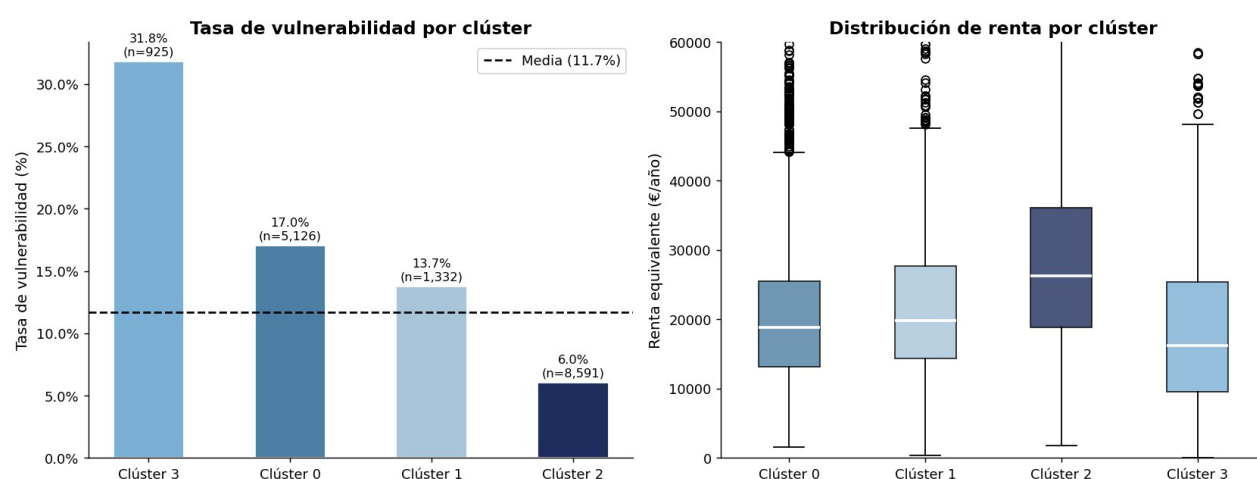


Figura 12. Tasa de vulnerabilidad (izquierda) y distribución de renta equivalente (derecha) por clúster.

6. Conclusiones

6.1. Síntesis de hallazgos principales

Este trabajo ha analizado la vulnerabilidad económica de los mayores de 65 años en España a partir de los microdatos de la ECV 2025, combinando un índice multidimensional, un modelo de regresión logística y un clustering k-means. Los hallazgos principales son los siguientes.

Primero, la tasa de vulnerabilidad financiera se sitúa en el 11,7%, equivalente a aproximadamente 1,1 millones de personas en la población total. Este resultado se obtiene aplicando el criterio de al menos dos dimensiones de fragilidad simultáneas, coherente con el enfoque multidimensional de Alkire y Foster (2011).

Segundo, la pobreza monetaria infravalora la fragilidad real del colectivo. La tensión presupuestaria mensual afecta al 32,8% y la ausencia de colchón financiero al 26,6%, ambas muy por encima del 16,4% con ingresos bajo el umbral. El 29,1% declara tensión mensual sin ser formalmente pobre, una desconexión que justifica el enfoque multidimensional adoptado.

Tercero, los tres factores con mayor efecto marginal sobre la probabilidad de ser vulnerable son la pobreza energética (+7,21 pp), la alta carga de vivienda (+6,84 pp) y el hogar unipersonal (+5,30 pp). La edad pierde significatividad una vez controladas estas variables, lo que confirma que la vulnerabilidad no es una consecuencia directa del envejecimiento biológico sino de configuraciones específicas de ingresos, vivienda y composición del hogar.

Cuarto, la educación es el factor protector más potente: -10,76 puntos porcentuales para los de educación alta respecto a los de educación baja (OR=0,272). Su efecto acumulado a lo largo de décadas de carrera laboral no se compensa con ningún otro factor individual.

Quinto, el modelo produce dos resultados aparentemente contraintuitivos: ser mujer (OR=0,573) y tener alquiler de mercado (OR=0,444) aparecen como factores protectores netos en el multivariante. Ambos se explican por efectos de composición: la desventaja del género opera a través de la pensión, ya controlada en el modelo, y los seniors en alquiler se concentran en zonas urbanas con pensiones más altas. Estos coeficientes no implican que mujeres o arrendatarios sean menos vulnerables en términos absolutos. De hecho, el Clúster

3, que combina alquiler e ingresos bajos, presenta la mayor tasa de vulnerabilidad del colectivo (31,8%).

Sexto, el clustering identifica cuatro perfiles con características y necesidades diferenciadas: Clúster 0 "Mujeres mayores solas" (17,0%), Clúster 1 "Dependencia funcional" (13,7%), Clúster 2 "Seniors protegidos" (6,0%) y Clúster 3 "Vulnerabilidad estructural por alquiler" (31,8%). Esta taxonomía tiene mayor utilidad operativa que los coeficientes del modelo para el diseño de intervenciones específicas.

6.1.1. Contraste de hipótesis

Las seis hipótesis formuladas en el punto 3.7 permiten evaluar la coherencia entre la teoría y la evidencia empírica.

H1 (mujeres → mayor riesgo): refutada en el modelo, confirmada en el descriptivo. Las mujeres tienen una tasa descriptiva del 13,2% frente al 9,9% de los hombres. Sin embargo, el modelo multivariante muestra $OR=0,57$: con la misma pensión y carrera laboral que un hombre, la mujer tiene menor probabilidad de ser vulnerable. La desventaja femenina opera a través de la pensión, no directamente por el sexo.

H2 (educación alta → menor riesgo): confirmada. $OR=0,27$ ($p<0,001$), el efecto protector más potente del modelo.

H3 (alquiler → mayor riesgo): refutada en el modelo, confirmada en el clustering. $OR=0,44$ en el multivariante por efecto de composición urbana. El Clúster 3 (alquiler + ingresos bajos) sí presenta la mayor tasa de vulnerabilidad (31,8%).

H4 (mala salud → mayor riesgo): confirmada. $OR>1$ con $p<0,05$, consistente en todos los niveles del análisis.

H5 (vivir solo → mayor riesgo): confirmada. $OR=1,90$ ($p<0,001$), uno de los efectos más potentes del modelo.

H6 (pobreza energética → mayor riesgo): confirmada. $OR=2,40$ ($p<0,001$), el odds ratio más alto del modelo.

Cuatro de seis hipótesis se confirman con solidez. Las dos aparentes refutaciones (H1 y H3) no son errores del análisis sino hallazgos: revelan que el sexo y el régimen de tenencia operan de forma indirecta, mediados por la pensión y la localización geográfica.

6.2. Implicaciones para políticas públicas

Los resultados señalan tres áreas prioritarias de intervención pública, cada una vinculada a uno de los principales factores de riesgo identificados.

La primera es la pobreza energética, que aparece como el predictor más potente del modelo (AME: +7,21 pp). Su elevada prevalencia entre los seniors vulnerables sugiere la necesidad de ampliar el bono social eléctrico con criterios específicos para pensionistas de rentas bajas y de reforzar los programas de rehabilitación energética en viviendas habitadas por mayores. La eficacia esperada es alta porque se actúa sobre un gasto recurrente que erosiona el margen presupuestario mes a mes.

La segunda es la vivienda en régimen de alquiler. El Clúster 3 concentra la mayor tasa de vulnerabilidad del colectivo (31,8%) y combina ingresos bajos con cuotas de alquiler que absorben dos tercios o más de la pensión. Las respuestas más directas pasan por ampliar el parque de vivienda asequible reservada a mayores, regular los incrementos de renta en contratos de inquilinos seniors y desarrollar ayudas al alquiler específicas para pensionistas, dado que las ayudas generales no siempre alcanzan a este perfil.

La tercera es la brecha de género en pensiones. La vulnerabilidad del Clúster 0 refleja las consecuencias acumuladas del modelo contributivo sobre quienes dedicaron su vida activa a tareas de cuidado no remuneradas. Medidas como el complemento por brecha de género, el reconocimiento de períodos de cuidado como cotización efectiva y la revalorización de las pensiones mínimas tienen un impacto directo sobre este perfil, que representa casi un tercio de la población senior.

6.3. Implicaciones para el sector asegurador

Los cuatro clústeres sugieren tres líneas de producto diferenciadas para MAPFRE y el sector asegurador en general.

Clúster 3 "Vulnerabilidad estructural por alquiler" (5,8% de la muestra, 31,8% de vulnerabilidad): protección de rentas básica. Productos que garanticen un ingreso mínimo ante imprevistos o que complementen la pensión hasta un nivel que permita afrontar el alquiler sin tensión presupuestaria. Aunque pequeño en tamaño, concentra el mayor riesgo y es el de mayor urgencia actuarial.

Clúster 0 "Mujeres mayores solas" (32,1% de la muestra, 17,0% de vulnerabilidad): coberturas adaptadas al perfil de pensión mínima. Rentas vitalicias con garantías de

supervivencia, seguros de vida que protejan ante la pérdida de la pensión del cónyuge y productos de ahorro-inversión con liquidez inmediata. Su tamaño, casi un tercio de la muestra senior, lo convierte en el segmento con mayor mercado potencial.

Clúster 1 "Dependencia funcional" (8,3% de la muestra, 13,7% de vulnerabilidad): seguros de dependencia y coberturas de cuidados de larga duración. La combinación de limitación funcional severa (99,6%), edad avanzada (79,2 años de media) y ausencia de colchón financiero (37,9%) genera una demanda clara de productos que cubran los costes que el SAAD no financia completamente.

Clúster 2 "Seniors protegidos" (53,8% de la muestra, 6,0% de vulnerabilidad): ahorro complementario y planificación patrimonial. Aunque el riesgo actual es bajo, su mayor capacidad de pago y numerosidad lo hacen el segmento más atractivo para seguros de dependencia preventivos contratados mientras la salud es buena.

6.4. Limitaciones del estudio

El análisis presenta cuatro limitaciones. Primera, la endogeneidad entre salud e ingresos (Deaton, 2003) impide interpretar los coeficientes como efectos causales. Segunda, la ECV no recoge activos financieros ni inmobiliarios, lo que puede generar sobreestimación de la vulnerabilidad en hogares con ahorro no declarado. Tercera, con el 11,7% de vulnerables, la sensibilidad del modelo con umbral 0,5 es baja (15,7%), lo que limita su uso para clasificación individual, aunque el AUC de 0,827 confirma buena capacidad de ordenamiento probabilístico y resulta válido para el análisis de factores de riesgo. Cuarta, la naturaleza transversal de los datos no permite analizar la dinámica temporal de la vulnerabilidad.

6.5. Líneas futuras de investigación

Las cuatro líneas más prometedoras son: combinar la ECV con la Encuesta Financiera de las Familias para incorporar la dimensión de patrimonio; utilizar el módulo longitudinal de la ECV para estudiar los shocks que desencadenan transiciones hacia la vulnerabilidad; extender el análisis a otros países europeos con distintos sistemas de pensiones y mercados de vivienda; y explorar modelos de árbol (random forest, gradient boosting) que capturan interacciones complejas y permiten comparación sistemática con el modelo logístico.

Referencias bibliográficas

- Abellán García, A., Aceituno, M. P., Pérez, J., Ramiro Fariñas, D., Ayala García, A., & Pujol Rodríguez, R. (2022). Un perfil de las personas mayores en España, 2022. Indicadores estadísticos básicos. CSIC. <https://doi.org/10.20350/digitalCSIC/14545>
- Alkire, S., & Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics*, 95(7-8), 476-487. <https://doi.org/10.1016/j.jpubeco.2010.11.006>
- Ando, A., & Modigliani, F. (1963). The "life cycle" hypothesis of saving: Aggregate implications and tests. *The American Economic Review*, 53(1), 55-84.
- Banco de España. (2020). Encuesta Financiera de las Familias (EFF) 2020: métodos, resultados y cambios desde 2017. Banco de España.
- Banco de España. (2022). Informe anual 2022. Banco de España. <https://www.bde.es/bde/es/secciones/informes/publicaciones-an/informe-anual/>
- Chaudhuri, S., Jalan, J., & Suryahadi, A. (2002). Assessing household vulnerability to poverty from cross-sectional data: A methodology and estimates from Indonesia (Discussion Paper 0102-52). Columbia University Department of Economics.
- Comas-Herrera, A., Wittenberg, R., Costa-Font, J., Gori, C., Di Maio, A., Patxot, C., Pickard, L., Pozzi, A., & Rothgang, H. (2006). Future long-term care expenditure in Germany, Spain, Italy and the United Kingdom. *Ageing & Society*, 26(2), 285-302. <https://doi.org/10.1017/S0144686X05004289>
- Comisión Europea y Comité de Protección Social. (2024). The 2024 pension adequacy report: Current and future income adequacy in old age in the EU. Oficina de Publicaciones de la Unión Europea.
- Deaton, A. (2003). Health, inequality, and economic development. *Journal of Economic Literature*, 41(1), 113-158. <https://doi.org/10.1257/002205103321544710>
- Eurostat. (2023). People at risk of poverty or social exclusion. Statistics Explained. European Commission. <https://ec.europa.eu/eurostat/statistics-explained/>
- Finkelstein, A., & McGarry, K. (2006). Multiple dimensions of private information: Evidence from the long-term care insurance market. *American Economic Review*, 96(4), 938-958.
- ICEA. (2022). Estadística de seguros de personas: informe anual 2022. Investigación Cooperativa entre Entidades Aseguradoras y Fondos de Pensiones.
- IMSERSO. (2022). Informe 2021: Las personas mayores en España. Datos estadísticos estatales y por comunidades autónomas. Instituto de Mayores y Servicios Sociales.

- Instituto Nacional de Estadística. (2023a). Encuesta de Condiciones de Vida (ECV) 2023. INE.
https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176807
- Instituto Nacional de Estadística. (2023b). Proyecciones de Población 2022-2072. INE.
https://www.ine.es/prensa/pp_2022_2072.pdf
- Instituto Nacional de Estadística. (2025). Encuesta de Condiciones de Vida (ECV) 2025: ficheros de microdatos. INE.
- Jalan, J., & Ravallion, M. (1998). Transient poverty in postreform rural China. *Journal of Comparative Economics*, 26(2), 338-357. <https://doi.org/10.1006/jcec.1998.1526>
- Ligon, E., & Schechter, L. (2003). Measuring vulnerability. *The Economic Journal*, 113(486), C95-C102. <https://doi.org/10.1111/1468-0297.00117>
- Lowsky, D. J., Olshansky, S. J., Bhattacharya, J., & Goldman, D. P. (2014). Heterogeneity in healthy aging. *The Journals of Gerontology: Series A*, 69(6), 640-649. <https://doi.org/10.1093/gerona/glt162>
- Lusardi, A., & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature*, 52(1), 5-44. <https://doi.org/10.1257/jel.52.1.5>
- Organización para la Cooperación y el Desarrollo Económicos. (2023). Pensions at a glance 2023: OECD and G20 indicators. OECD Publishing. <https://doi.org/10.1787/678d8133-en>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
- Schröder-Butterfill, E., & Marianti, R. (2006). A framework for understanding old-age vulnerabilities. *Ageing & Society*, 26(1), 9-35. <https://doi.org/10.1017/S0144686X05004423>
- Smith, J. P. (1999). Healthy bodies and thick wallets: The dual relation between health and economic status. *Journal of Economic Perspectives*, 13(2), 145-166. <https://doi.org/10.1257/jep.13.2.145>

Declaración de uso de herramientas de IAG

Por la presente, yo, Teresa Suquía Vega, estudiante de 5º de E2 + Analytics B de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado "Understanding company valuation: A practical approach applied to Inditex", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. **Brainstorming de ideas de investigación:** Utilizado para idear y esbozar posibles áreas de investigación.
2. **Referencias:** Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
3. **Constructor de plantillas:** Para diseñar formatos específicos para secciones del trabajo.
4. **Corrector de estilo literario y de lenguaje:** Para mejorar la calidad lingüística y estilística del texto.
5. **Generador previo de diagramas de flujo y contenido:** Para esbozar diagramas iniciales.
6. **Sintetizador y divulgador de libros complicados:** Para resumir y comprender literatura compleja.
7. **Revisor:** Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 16/06/2026

Firma: _____

