



ICADE / E2 + Business Analytics

ALFABETIZACIÓN FINANCIERA Y DECISIONES DE AHORRO E INVERSIÓN DE LOS JÓVENES EN ESPAÑA

Autor: Lucas Esnal Olarreaga

Tutor: Raquel Redondo Palomo

MADRID | Junio de 2026

RESUMEN

Este trabajo analiza la relación entre la alfabetización financiera de los jóvenes españoles y sus decisiones de ahorro e inversión. Para ello emplea los microdatos de la Encuesta de Competencias Financieras de 2021, elaborada por el Banco de España y la Comisión Nacional del Mercado de Valores, sobre una muestra de 1.899 jóvenes de 18 a 34 años. El conocimiento financiero se mide a través del índice de las *Big Three* y se relaciona con el comportamiento mediante modelos de regresión lineal y logística, complementados con una segmentación en perfiles. Los resultados muestran que el nivel de conocimiento de los jóvenes es algo inferior al del conjunto de la población y que la formación académica es su principal determinante, junto a una brecha de género que se mantiene entre individuos comparables. El hallazgo central es que la alfabetización financiera apenas se relaciona con la decisión de ahorrar, pero sí con la de invertir. El conocimiento no determina si un joven ahorra, sino la forma en que decide emplear lo ahorrado. La segmentación confirma este patrón, ya que la formación distingue al grupo inversor, mientras que la capacidad de ahorro separa a quienes ahorran de quienes no lo consiguen. Estos resultados sugieren que la educación financiera dirigida a los jóvenes resultaría más eficaz si se orientara al paso del ahorro a la inversión, más que a fomentar el ahorro en sí mismo.

Palabras clave: alfabetización financiera, jóvenes, ahorro, inversión, Encuesta de Competencias Financieras, España.

ABSTRACT

This study analyses the relationship between the financial literacy of young people in Spain and their saving and investment decisions. It draws on microdata from the 2021 Survey of Financial Competences, conducted by the Banco de España and the Comisión Nacional del Mercado de Valores, covering a sample of 1,899 individuals aged 18 to 34. Financial literacy is measured through the Big Three index and linked to behaviour using linear and logistic regression models, complemented by a cluster-based segmentation. The results show that young people's level of knowledge is somewhat lower than that of the overall population and that educational attainment is its main determinant, alongside a gender gap that persists among comparable individuals. The central finding is that financial literacy is barely related to the decision to save, but it is related to the decision to invest. Knowledge does not determine whether a young person saves, but rather how they choose to allocate what they have saved.

The segmentation confirms this pattern, as education sets the investor group apart, while saving capacity separates those who manage to save from those who do not. These findings suggest that financial education aimed at young people would be more effective if it focused on the transition from saving to investing, rather than on encouraging saving itself.

Keywords: financial literacy, youth, saving, investment, Survey of Financial Competences, Spain.

Índice

1. INTRODUCCIÓN	6
1.1. Justificación y motivación	6
1.2. Objetivos	8
1.3. Estructura del trabajo	8
2. MARCO TEÓRICO.....	10
2.1. Concepto, medición y relevancia de la alfabetización financiera	10
2.1.1. Concepto	10
2.1.2. Medición: las <i>Big Three</i>	10
2.1.3. Relevancia.....	11
2.2. Determinantes de la alfabetización financiera	11
2.3. Alfabetización financiera, ahorro e inversión	13
2.4. La alfabetización financiera de los jóvenes en España	13
2.4.1. Un contexto socioeconómico exigente	13
2.4.2. Diferencias dentro del grupo.....	14
2.4.3. Formación y digitalización	15
2.5. Hipótesis de investigación	15
3. DATOS Y METODOLOGÍA.....	16
3.1. La Encuesta de Competencias Financieras de 2021	16
3.2. Muestra y definición de la cohorte joven.....	16
3.3. Variables del análisis	18
3.4. Técnicas de análisis.....	19
4. ANÁLISIS Y RESULTADOS	21
4.1. Nivel y distribución de la alfabetización financiera	21
4.2. Determinantes del conocimiento financiero de los jóvenes.....	22
4.3. Alfabetización y tenencia de productos de ahorro e inversión	23
4.4. Perfiles de jóvenes según alfabetización y comportamiento	24
4.5. Análisis de robustez	26
5. DISCUSIÓN	28
5.1. Interpretación de los resultados	28
5.2. Implicaciones	29
5.3. Limitaciones.....	30
6. CONCLUSIONES	31
6.1. Síntesis de los resultados	31
6.2. Respuesta a las hipótesis.....	31
6.3. Aportaciones y líneas futuras.....	32

Bibliografía	34
Anexo 1	36
Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado.....	36

1. INTRODUCCIÓN

1.1. Justificación y motivación

Las decisiones financieras forman parte de la vida cotidiana. Abrir una cuenta, pedir un préstamo, contratar un seguro o destinar una parte del salario al ahorro son elecciones que casi todas las personas afrontan en algún momento. La forma en que se resuelven influye en el bienestar económico a lo largo de los años. Por ese motivo, la capacidad de cada individuo para tomar decisiones financieras informadas ha pasado a ocupar un lugar central en el debate económico y social.

Esa capacidad ha ganado importancia por un cambio de fondo en el reparto de responsabilidades. Durante décadas, buena parte de las decisiones económicas que afectaban al futuro de las familias recaía en instituciones como el Estado, las empresas o las entidades financieras. En los últimos años, esa responsabilidad se ha trasladado de manera creciente hacia el propio individuo. El ahorro para la jubilación es el caso más visible, pues los sistemas de pensiones de muchos países han evolucionado hacia esquemas en los que el resultado final depende de las aportaciones y de las decisiones de inversión de cada persona (Lusardi y Mitchell, 2014).

A ese cambio se suma una oferta financiera cada vez más amplia y compleja, en la que junto a los instrumentos tradicionales han aparecido otros más sofisticados y la digitalización ha multiplicado las vías de acceso. Esta abundancia ofrece más posibilidades, pero también exige más criterio, ya que la dificultad de las decisiones aumenta a medida que se amplía el abanico de opciones disponibles (Lusardi, 2019). Quien no domina los conceptos básicos queda en desventaja frente a un entorno que da por supuesta una competencia que no todos poseen.

En este contexto, los jóvenes constituyen un grupo de especial interés. Se encuentran en el momento del ciclo vital en el que se toman las primeras decisiones económicas de cierta importancia y en el que empiezan a formarse hábitos que tienden a mantenerse con el tiempo (Lusardi, Mitchell y Curto, 2010). Las elecciones que realizan en esta etapa, sobre el ahorro, el endeudamiento o la inversión, condicionan en buena medida su situación futura. Por eso, comprender cómo se enfrentan a esas decisiones resulta particularmente relevante.

La situación de los jóvenes en España añade motivos para fijar en ellos la atención. Las generaciones recientes afrontan un panorama económico más exigente que el de sus predecesores. La emancipación se retrasa, el acceso a la vivienda se ha encarecido y la dependencia del hogar familiar se prolonga más allá de lo que era habitual (Consejo de la Juventud de España, 2026). A estas dificultades materiales se une un punto de partida en materia de conocimientos financieros que deja margen de mejora. La combinación de ambos factores convierte a este grupo en un objeto de estudio pertinente, porque es precisamente donde las decisiones económicas resultan más determinantes y donde una preparación insuficiente puede tener consecuencias más duraderas.

El interés por el conocimiento financiero no es solo individual, sino también colectivo. Una población mejor preparada para gestionar sus recursos tiende a tomar decisiones más sólidas, lo que se traduce en una mayor estabilidad económica del conjunto. Por esta razón, numerosos organismos internacionales han impulsado estrategias nacionales dirigidas a mejorar la formación financiera de la ciudadanía, una línea de actuación respaldada por el G20 (OCDE/INFE, 2012). España se ha sumado a ese esfuerzo, lo que ha generado la necesidad de disponer de información fiable sobre el nivel real de conocimientos de la población y sobre los factores que lo explican. Sin un diagnóstico preciso, resulta difícil diseñar medidas eficaces.

Conviene, además, distinguir entre saber y hacer. No basta con medir cuánto conocen las personas sobre cuestiones económicas, sino que importa también averiguar si ese conocimiento se traduce en comportamientos concretos. Una persona puede comprender bien los conceptos básicos y, pese a ello, no llevar sus decisiones en consonancia con ese saber. Esa distancia resulta especialmente interesante en el caso del ahorro y la inversión, dos decisiones que no responden a la misma lógica ni exigen el mismo grado de competencia. La Encuesta de Competencias Financieras, elaborada por el Banco de España y la Comisión Nacional del Mercado de Valores, permite abordar estas cuestiones con rigor, ya que ofrece información detallada sobre los conocimientos económicos de la población y sobre la tenencia y el uso de productos financieros (Banco de España, 2023).

1.2. Objetivos

El objetivo general de este trabajo es analizar la relación entre la alfabetización financiera de los jóvenes en España y sus decisiones de ahorro e inversión, a partir de los microdatos de la Encuesta de Competencias Financieras de 2021.

Para alcanzar ese propósito general, se plantean cuatro objetivos específicos.

- Describir el nivel de conocimiento financiero de los jóvenes españoles y situarlo en relación con el del conjunto de la población.
- Identificar las características personales asociadas a las diferencias de conocimiento dentro del grupo.
- Determinar en qué medida el conocimiento financiero se relaciona con la tenencia de productos de ahorro e inversión, distinguiendo la decisión de ahorrar de la de invertir.
- Agrupar a los jóvenes en perfiles diferenciados según su nivel de conocimiento y su comportamiento financiero.

El primer objetivo tiene un carácter descriptivo y sirve de punto de partida. Los dos siguientes constituyen el núcleo del trabajo, ya que abordan tanto las causas de las diferencias de conocimiento como su conexión con la conducta. El último ofrece una mirada complementaria, orientada a reconocer la diversidad interna del grupo. En conjunto, los cuatro permiten pasar de una visión general del nivel de conocimiento a un análisis más detallado de su relación con las decisiones económicas.

1.3. Estructura del trabajo

El trabajo se organiza en seis capítulos, incluida esta introducción, que avanzan desde el planteamiento conceptual hasta las conclusiones siguiendo un orden progresivo.

El segundo capítulo presenta el marco teórico. En él se revisa el concepto de alfabetización financiera, su medición a través de las preguntas que se han convertido en estándar internacional y su relevancia económica y social. A continuación, se examinan los factores que la determinan y se analiza su relación con el ahorro y la inversión. El capítulo dedica un apartado específico a la situación de los jóvenes en España y concluye con la formulación de las hipótesis que guían el análisis.

El tercer capítulo describe los datos y la metodología. Detalla las características de la Encuesta de Competencias Financieras y los motivos de su elección, define la muestra de jóvenes objeto de estudio y su división en tramos de edad, presenta las variables empleadas y justifica las técnicas estadísticas utilizadas, que combinan el análisis descriptivo, los modelos de regresión y la segmentación en grupos.

El cuarto capítulo recoge el análisis y los resultados. Comienza con una descripción del nivel de conocimiento de la cohorte y su comparación con el resto de la población, continúa con el estudio de los factores que explican las diferencias internas y de la relación entre conocimiento y tenencia de productos, y concluye con una segmentación en perfiles y una comprobación de la solidez de los resultados frente a decisiones metodológicas alternativas.

El quinto capítulo discute los hallazgos. Interpreta su significado más allá de las cifras, los pone en relación con la evidencia previa, extrae implicaciones para el diseño de la educación financiera y reconoce las limitaciones que conviene tener presentes al valorar las conclusiones.

El sexto y último capítulo sintetiza los principales resultados, responde de forma explícita a las hipótesis planteadas en el marco teórico y señala tanto las aportaciones del trabajo como las posibles líneas de investigación futura.

2. MARCO TEÓRICO

Las decisiones financieras dependen, entre otros factores, de los conocimientos con los que cada persona interpreta la información económica y orienta el uso de su dinero. Cuánto domina la población esas cuestiones, y en qué medida ese dominio se refleja en su conducta, son dos de las preguntas que más atención han recibido en la economía de las últimas décadas. Esa competencia recibe el nombre de alfabetización financiera y es el objeto central de este trabajo.

2.1. Concepto, medición y relevancia de la alfabetización financiera

2.1.1. Concepto

La alfabetización financiera se ha consolidado como uno de los conceptos de referencia en el estudio del comportamiento financiero individual. Lusardi y Mitchell (2014) la definen como «la capacidad de las personas para procesar información económica y tomar decisiones informadas sobre planificación financiera, acumulación de riqueza, deuda y pensiones». Su aportación clave está en subrayar que el saber solo cobra valor cuando llega a la práctica; una persona puede dominar la teoría y, aun así, no aplicarla correctamente a sus decisiones económicas (Huston, 2010).

Esta competencia presenta además dos caras que no siempre coinciden entre sí. Una cosa es el conocimiento que una persona posee en realidad y otra el que cree tener. Cuando existe una discrepancia entre ambos, quienes tienen menor nivel de conocimiento tienden con más frecuencia a sobrestimar su propia capacidad, lo que puede llevarlos a tomar decisiones por encima de su criterio real (Allgood y Walstad, 2016).

2.1.2. Medición: las *Big Three*

Para medir esa capacidad, Lusardi y Mitchell (2011) idearon tres preguntas sobre la inflación, el interés compuesto y la diversificación del riesgo. Conocidas como las *Big Three*, estas cuestiones se han convertido en el estándar internacional de referencia y se han incorporado a encuestas nacionales de numerosos países (Lusardi, 2019). Entre ellas figura la Encuesta de Competencias Financieras (ECF) que el Banco de España aplica desde 2016 (Banco de España, 2023).

La elección de esos tres conceptos no es casual, pues cada uno resulta básico para casi cualquier decisión económica (Lusardi, 2019). El interés compuesto explica cómo crece el dinero con el tiempo, lo que lo hace esencial para entender tanto el ahorro como el coste de la deuda. La inflación recuerda que los precios suben y que unos ahorros parados pierden valor. La diversificación del riesgo recoge la idea de que repartir una inversión entre varios activos reduce la exposición a pérdidas. Son nociones sencillas en apariencia, pero su comprensión conjunta marca la frontera entre un manejo solvente de las finanzas y otro más vulnerable.

Conviene además situar esta aptitud en su contexto actual. La digitalización ha multiplicado el acceso a productos y datos financieros, pero no la habilidad para interpretarlos. Esa distancia crece conforme la oferta se vuelve cada vez más compleja (CNMV y Banco de España, 2022). La alfabetización financiera deja así de ser una cualidad deseable para convertirse en una herramienta necesaria, lo que explica el interés por analizarla y por comprender sus determinantes.

2.1.3. Relevancia

El interés por medir esta competencia responde, en última instancia, a sus consecuencias prácticas. Las personas con mayor conocimiento financiero planifican con más frecuencia su jubilación y acumulan más riqueza a lo largo de la vida (Lusardi y Mitchell, 2014). De las tres dimensiones básicas, la comprensión del riesgo es la que más se relaciona con esa previsión a largo plazo.

Las carencias, en cambio, tienen un coste. Una baja alfabetización financiera se asocia con decisiones económicas menos acertadas, un riesgo que se agrava a medida que la oferta gana en complejidad y exige al consumidor elecciones cada vez más informadas (Lusardi, 2019). Por su impacto sobre el bienestar individual, su mejora se ha convertido en un objetivo recurrente de las políticas públicas.

2.2. Determinantes de la alfabetización financiera

El conocimiento financiero no se reparte de manera uniforme y las diferencias observadas siguen patrones similares incluso entre países con sistemas educativos distintos (Lusardi y Mitchell, 2014). Identificar los factores asociados a estas desigualdades permite

localizar a los colectivos con mayores carencias e interpretar mejor la relación entre lo que se sabe y lo que se hace. Estos determinantes se suelen agrupar en tres bloques:

- **Factores socioeconómicos:** la educación es el factor más influyente. Aquellos con estudios superiores responden bien a las preguntas básicas con mucha más frecuencia que quienes únicamente completaron la enseñanza primaria (Mancebón et al., 2020).

La renta apunta en la misma dirección, aunque su lectura es menos directa. Disponer de mayores recursos facilita el acceso a productos financieros. La relación, además, podría operar también a la inversa, ya que quien dispone de más dinero tiene más ocasiones e incentivos para adquirir esos conocimientos (Lusardi y Mitchell, 2014).

- **Factores demográficos:** el conocimiento financiero no evoluciona de forma lineal a lo largo de la vida. Sigue un patrón de U invertida con su nivel más bajo en la juventud, el máximo en las edades centrales y un nuevo descenso en la vejez (Mancebón et al., 2020).

En cuanto al género, los hombres superan a las mujeres en la mayoría de países. Se trata de una de las diferencias más constantes en este campo, aunque sus causas no se conocen del todo. Una explicación habitual apunta al reparto tradicional de tareas dentro del hogar, donde la gestión financiera ha recaído más a menudo en los hombres, lo que les ha dado mayor contacto con estos asuntos (Hsu, 2011).

- **Factores familiares y laborales:** el hogar de origen influye a través del aprendizaje por observación. Crecer donde se habla de dinero y se manejan productos financieros se asocia con un mayor conocimiento de los hijos, hasta el punto de que la educación de los padres y la tenencia de activos en la familia figuran entre los indicadores más relacionados con el saber juvenil (Lusardi et al., 2010).

Posteriormente, la incorporación al empleo aporta una fuente adicional de aprendizaje práctico, ya que obliga a administrar un sueldo y a decidir qué hacer con él.

Ninguno de estos factores actúa de forma aislada. Por el contrario, suelen interactuar entre sí y dar lugar a trayectorias financieras diferentes. Distinguir el efecto real de cada uno exige, por tanto, considerarlos a la vez.

2.3. Alfabetización financiera, ahorro e inversión

El interés por medir la alfabetización nace, en última instancia, de su relación con el comportamiento. Entre los adultos, quienes puntúan alto en las preguntas básicas tienen más probabilidad de disponer de un colchón de emergencia y menos de recurrir a créditos de coste elevado (de Bassa Scheresberg, 2013). El hallazgo se ha replicado en numerosos contextos y constituye el principal argumento a favor de las políticas de educación financiera.

Sin embargo, no todas estas elecciones plantean las mismas exigencias. Conviene entonces distinguir entre ahorro e inversión, dos conceptos estrechamente relacionados pero que responden a lógicas diferentes. El ahorro es la parte de la renta que no se destina al consumo y depende sobre todo de los ingresos disponibles y de los gastos comprometidos. Invertir va más allá, pues consiste en asignar esos recursos a activos que ofrecen una rentabilidad esperada a cambio de asumir cierto riesgo. Exige, por ello, comparar alternativas y comprender productos que a veces resultan complejos. Cabe esperar, por tanto, que el conocimiento influya poco en la decisión inicial de ahorrar y bastante más a la hora de decidir cómo emplear ese ahorro (van Rooij et al., 2011).

Esta relación, en cualquier caso, debe interpretarse con cautela. Que los más formados inviertan más no implica que una cosa cause la otra, ya que factores como la capacidad numérica o las preferencias personales pueden influir sobre ambas (Fernandes et al., 2014). De hecho, los estudios que tratan de aislar ese efecto encuentran resultados más moderados que las correlaciones simples y confirman que la formación mejora la conducta solo de forma limitada (Kaiser et al., 2022).

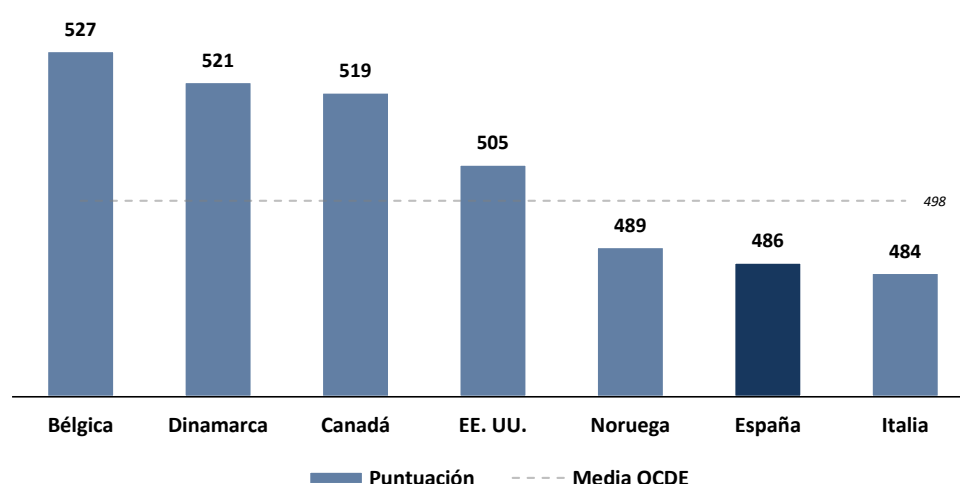
2.4. La alfabetización financiera de los jóvenes en España

2.4.1. Un contexto socioeconómico exigente

Los jóvenes españoles reúnen condiciones que hacen especialmente oportuno su estudio. Atraviesan la etapa en la que se contraen los primeros compromisos económicos y se consolidan hábitos llamados a perdurar. Lo hacen, además, en un contexto material poco favorable, con una emancipación cada vez más tardía y un alquiler que absorbe gran parte del salario medio de un joven trabajador (Consejo de la Juventud de España, 2026). Como consecuencia, las generaciones recientes acumulan menos patrimonio que las anteriores a la misma edad (Martínez Santos y Galindo, 2025).

A esas restricciones se suma un nivel de conocimiento mejorable. En la ECF, los menores de 35 años figuran entre los grupos con menos aciertos (Banco de España, 2023). El déficit se aprecia incluso antes de la vida adulta, ya que los estudiantes españoles obtuvieron 486 puntos en PISA 2022, por debajo del promedio de la OCDE, situado en 498. Uno de cada seis alumnos no alcanzó el nivel mínimo y apenas un 5 % llegó al más alto (OCDE, 2024). Esa puntuación, además, supuso un retroceso de seis puntos frente a 2018, en contraste con la mejora registrada en la mayoría de los países participantes (OCDE, 2024).

Gráfica 1. Puntuación en Competencia Financiera



Fuente: elaboración propia a partir de PISA (2022)

2.4.2. Diferencias dentro del grupo

Detrás de esa media conviven, sin embargo, perfiles muy distintos. Entre los adolescentes, el conocimiento varía con el tamaño del municipio y el origen del hogar, de modo que las zonas rurales y las familias inmigrantes parten en desventaja (Funcas, 2024). El entorno familiar refuerza esa desigualdad, ya que los estudiantes que hablan de dinero con sus padres obtienen mejores resultados que quienes no lo hacen (OCDE, 2024). Las diferencias entre chicas y chicos, en cambio, son reducidas a esas edades, lo que sugiere que la brecha de género se abre más tarde (Funcas, 2024).

El momento vital añade otra frontera. Entre los 18 y los 24 años predominan quienes estudian y dependen del hogar familiar, mientras que entre los 25 y los 34 el empleo y las primeras decisiones económicas multiplican las ocasiones de adquirir experiencia financiera. Esta diferencia aconseja analizar ambos tramos por separado.

2.4.3. Formación y digitalización

La evidencia disponible indica además que los jóvenes responden positivamente a la formación financiera. Un curso breve impartido en institutos españoles mejoró tanto el conocimiento del alumnado como su paciencia en decisiones de ahorro (Bover et al., 2018). Esta respuesta no es exclusiva del caso español, pues la evidencia internacional confirma que los programas educativos elevan el conocimiento financiero, con efectos especialmente marcados entre niños y jóvenes (Kaiser et al., 2022).

El último rasgo distintivo es la digitalización. Nueve de cada diez jóvenes conocen las criptomonedas (Banco de España, 2023). Aunque su tenencia seguía siendo minoritaria en 2021, quienes las poseen suelen presentar niveles relativamente elevados de formación y conocimientos financieros (Guirola e Isla Martínez, 2025). Se trata además de un perfil tolerante al riesgo, que muchas veces convive aún en el hogar familiar y que no rechaza el sistema financiero tradicional. Aunque la exposición general es baja, una pequeña fracción concentra en estos activos una parte elevada de su patrimonio, lo que la hace más vulnerable a sus oscilaciones (Guirola e Isla Martínez, 2025).

2.5. Hipótesis de investigación

El recorrido anterior conecta el conocimiento financiero con la conducta económica y sitúa a los jóvenes españoles como un grupo donde esa relación resulta especialmente relevante. Sobre esa base se formulan las dos hipótesis que guían el análisis empírico:

- **H1:** entre los jóvenes españoles, un mayor nivel de alfabetización financiera se asocia con una probabilidad más alta de poseer productos de ahorro e inversión, una vez consideradas sus características sociodemográficas y económicas.
- **H2:** esa asociación es más intensa en la elección de instrumentos concretos que en la propia decisión de ahorrar.

Ambas hipótesis recogen el hilo argumental desarrollado a lo largo del capítulo. La primera traslada al terreno empírico la relación general entre conocimiento y conducta, mientras que la segunda incorpora el matiz central de toda la discusión, que esa relación no pesa igual en la decisión de ahorrar que en la de invertir.

3. DATOS Y METODOLOGÍA

3.1. La Encuesta de Competencias Financieras de 2021

La parte empírica se apoya en los microdatos de la Encuesta de Competencias Financieras de 2021, elaborada por el Banco de España y la Comisión Nacional del Mercado de Valores. La encuesta forma parte de una iniciativa internacional coordinada por la OCDE, lo que garantiza la comparabilidad de sus preguntas entre países. Su objetivo es medir el conocimiento de conceptos financieros básicos junto a la tenencia y el uso de productos de ahorro e inversión entre la población de 18 a 79 años residente en España.

El cuestionario va más allá de las tres preguntas básicas. Recoge la tenencia, la adquisición y el uso de un amplio conjunto de productos de ahorro, inversión y endeudamiento. Además, cuenta con información detallada sobre las características personales y económicas de cada hogar, lo que permite enlazar el conocimiento financiero con la conducta efectiva (Banco de España, 2023).

Este estudio emplea la segunda oleada, cuyas entrevistas se realizaron entre finales de 2021 y mediados de 2022, con resultados publicados en noviembre de 2023 (Banco de España, 2023). Frente a la primera edición, de 2016, la más reciente ofrece una imagen actualizada de la población e incorpora productos antes ausentes, entre ellos las criptomonedas, de especial interés al estudiar a los más jóvenes.

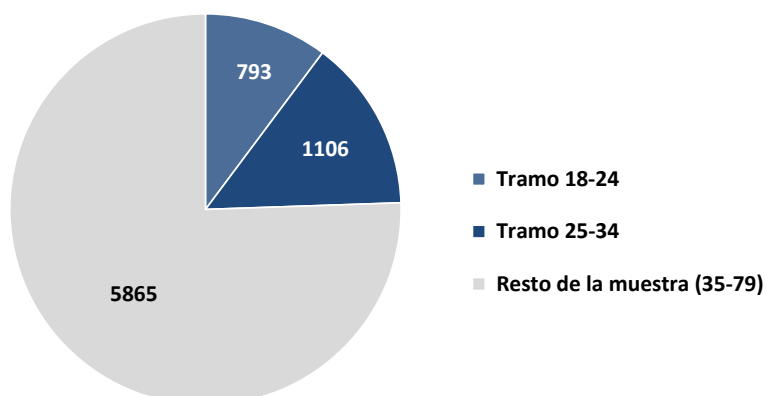
Las variables se han construido siguiendo los mismos criterios que el Banco de España, de modo que los resultados son directamente comparables con los publicados. La edición de 2016 se emplea solo en alguna comparación puntual, a partir de cifras ya difundidas (Banco de España, 2023). Los modelos principales se estiman sin los pesos muestrales, una opción habitual en este tipo de análisis. Su efecto se examina en el apartado de robustez, lo que explica que algún descriptivo difiera levemente de las cifras oficiales.

3.2. Muestra y definición de la cohorte joven

La encuesta recoge 7.764 entrevistas. De ese total, la cohorte de interés la forman los jóvenes de 18 a 34 años, lo que deja 1.899 observaciones, mientras que el resto de la muestra se emplea solo como término de comparación. Dentro del grupo se distinguen dos tramos, el de 18 a 24 años, con 793 individuos, y el de 25 a 34, con 1.106. Esta división permite comprobar

si la relación entre conocimiento y conducta cambia conforme avanza la transición a la vida adulta.

Gráfica 2. Distribución de la muestra por tramos de edad



Fuente: elaboración propia a partir de la Encuesta de Competencias Financieras 2021 (Banco de España y CNMV)

La Tabla 1 resume el perfil sociodemográfico de la cohorte. Las mujeres representan el 48 % y la edad media se sitúa en 26,1 años. En cuanto al nivel educativo, aproximadamente el 44 % ha alcanzado estudios universitarios, frente a un 37,7 % con estudios secundarios y un 18,4 % con estudios primarios. En el ámbito laboral, cerca de la mitad son asalariados y un 13 % se encuentra en el paro. El dato sobre vivienda requiere una interpretación cuidadosa; aunque el 64,4 % reside en un hogar en propiedad, ello no implica que estos jóvenes sean propietarios, ya que a estas edades muchos todavía viven en el hogar familiar. En realidad, la variable refleja el régimen de tenencia del hogar y no la propiedad individual.

Tabla 1. Perfil sociodemográfico de la cohorte joven (18-34 años)

Indicador	Valor
Mujeres	48,0%
Edad media	26,1 años
Estudios universitarios	43,9%
Estudios secundarios	37,7%
Estudios primarios	18,4%
Asalariados	50,8%
Autónomos	6,4%
Desempleados	13,0%
Jubilados y otros inactivos	29,9%
Vive en hogar con vivienda en propiedad	64,4%

Fuente: elaboración propia a partir de la Encuesta de Competencias Financieras 2021 (Banco de España y CNMV)

3.3. Variables del análisis

El análisis descansa sobre tres tipos de variables:

- **Alfabetización financiera:** la variable central es el índice *Big Three* descrito en el capítulo anterior, que suma un punto por cada una de las tres preguntas acertadas. Las respuestas «no sabe» o «no contesta» computan como fallo, si bien el efecto de tratarlas de otro modo se valora entre las comprobaciones de robustez. De forma complementaria se define una variable binaria de alfabetización plena, que vale uno cuando se aciertan las tres.
- **Conducta financiera:** el comportamiento de ahorro e inversión se capta de dos formas:
 - **Tenencia de productos:** codificada como sí o no para cada uno de los seis considerados: cuenta de ahorro, plan de pensiones, fondo de inversión, acciones, renta fija y criptomonedas. A partir de ellos se construyen dos variables; la primera indica si la persona dispone de al menos un producto, la segunda cuenta cuántos posee, de cero a seis.
 - **Hábito de ahorro:** recoge si la persona declara haber ahorrado en los últimos doce meses, con independencia de cómo se materialice ese ahorro.
- **Variables de control:** características sociodemográficas y económicas vinculadas al conocimiento financiero. Son cinco: el sexo, el tramo de edad, el nivel educativo, la situación laboral y la riqueza inmobiliaria del hogar, esta última como aproximación a sus recursos.

La renta del hogar merece una mención aparte. El modelo no la incluye entre las variables de control, ya que la encuesta no ofrece la cifra declarada por cada hogar, sino una estimación obtenida mediante técnicas de imputación. Su uso correcto requiere un tratamiento estadístico que excede el alcance de este trabajo y añadiría la incertidumbre propia de los datos estimados. En su lugar, la situación económica se aproxima a través del nivel educativo, la situación laboral y la riqueza inmobiliaria del hogar, tres variables estrechamente ligadas a la renta.

Tabla 2. Descripción de las variables del análisis

Variable	Definición y construcción	Categorías o valores
Índice de alfabetización (<i>Big Three</i>)	Suma de aciertos en las preguntas de inflación, interés compuesto y diversificación del riesgo; «no sabe» y «no contesta» computan como fallo	0 a 3
Alfabetización plena	Variable binaria derivada del índice	1 = tres aciertos; 0 = resto
Tenencia de cada producto	Indicador binario para cuenta de ahorro, plan de pensiones, fondo de inversión, acciones, renta fija y criptomonedas	1 = sí; 0 = no
Algún producto	Indicador agregado de tenencia	1 = al menos un producto; 0 = ninguno
Número de productos	Recuento de productos en cartera	0 a 6
Hábito de ahorro	Haber ahorrado en los últimos doce meses, con independencia del instrumento	1 = sí; 0 = no
Sexo	Control sociodemográfico	Mujer (ref.: hombre)
Tramo de edad	Control sociodemográfico	25-34 (ref.: 18-24)
Nivel educativo	Control sociodemográfico	Secundaria y universitaria (ref.: primaria)
Situación laboral	Control sociodemográfico	Asalariado, desempleado, jubilado y otros inactivos (ref.: autónomo)
Riqueza inmobiliaria del hogar	Aproximación a los recursos según el régimen de tenencia de la vivienda	No propietario; propietario solo de la vivienda principal; propietario de la principal y otras (ref.: otros regímenes o sin información)

Fuente: elaboración propia a partir de la Encuesta de Competencias Financieras 2021 (Banco de España y CNMV).

3.4. Técnicas de análisis

El análisis comienza con una exploración descriptiva y se apoya después en tres técnicas estadísticas.

- **Análisis descriptivo:** retrata a la cohorte, sitúa su conocimiento frente al del resto de la población y detecta patrones preliminares que orientan los modelos posteriores.
- **Regresión lineal (MCO):** modela el índice de alfabetización en función de las variables de control. Cada coeficiente se interpreta como la variación media en el número de aciertos respecto a la categoría de referencia, manteniendo constante el resto de factores. Si bien es cierto que existen modelos más específicos para este tipo de variable, pues el índice solo toma cuatro valores enteros, se opta por la regresión lineal porque sus coeficientes resultan más fáciles de interpretar y permiten comparar los resultados con los de otros análisis de la misma encuesta (Ispuerto et al., 2021).

Los resultados principales se estiman sin ponderar, dado que el objetivo es analizar las relaciones entre variables dentro de la muestra y facilitar la interpretación de los modelos. No obstante, para comprobar que las conclusiones no dependen de esta decisión, se replica el análisis utilizando los pesos muestrales de la ECF en el apartado de robustez.

- **Regresión logística (logit):** para cada conducta binaria se estima un modelo con el índice de alfabetización como variable explicativa principal y los mismos controles, de modo que conecta el conocimiento con cada decisión financiera concreta. Sus resultados se expresan como *odds ratios*, que resumen si cada acierto adicional hace más o menos probable la conducta. Un valor por encima de uno indica asociación positiva y un valor por debajo, negativa. Cada estimación se acompaña de su intervalo de confianza al 95 %.
- **Segmentación k-means:** en lugar de medir relaciones entre variables, agrupa a los individuos en perfiles homogéneos a partir de cuatro: conocimiento, edad, número de productos y hábito de ahorro reciente. A diferencia de las técnicas anteriores, no parte de una relación previa, sino que busca perfiles que estas no revelan. El número de grupos se determina mediante el coeficiente de silueta, que orienta sobre la partición más adecuada.

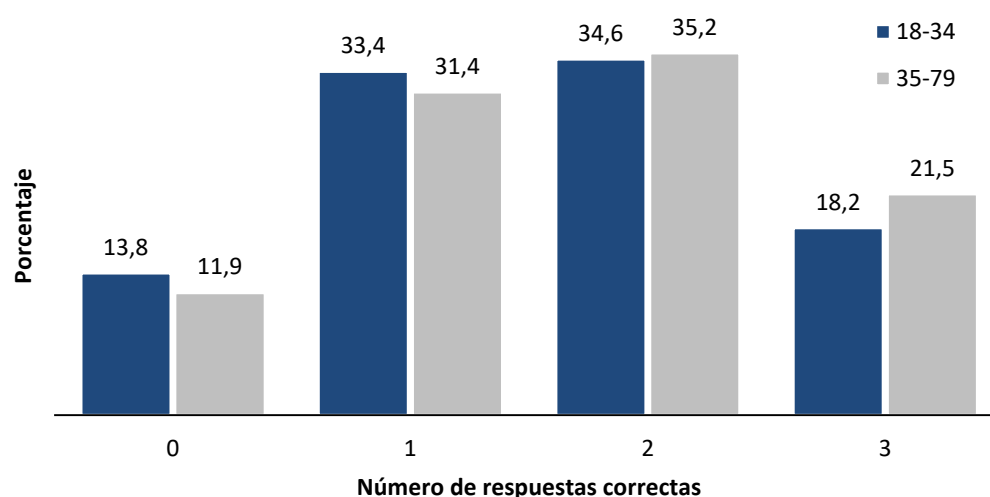
Por último, la estabilidad de los resultados se contrasta con tres comprobaciones, que se desarrollan junto a sus cifras más adelante.

4. ANÁLISIS Y RESULTADOS

4.1. Nivel y distribución de la alfabetización financiera

El conocimiento financiero de la cohorte se concentra en los niveles intermedios de la escala. La mayor parte del grupo se sitúa en uno o dos aciertos, con un 33,4 % y un 34,6 % respectivamente. El 18,2 % responde correctamente a las tres preguntas, frente al 20,7 % del total de encuestados. La comparación con la población de 35 años en adelante muestra una distribución ligeramente desplazada hacia la izquierda, ya que entre los mayores el grupo de tres aciertos sube al 21,5 %.

Gráfica 3. Distribución del índice de alfabetización: jóvenes frente al resto de la población



Fuente: elaboración propia a partir de la Encuesta de Competencias Financieras 2021 (Banco de España y CNMV)

La relación entre conocimiento y conducta se observa en los datos de la Tabla 3. Haber ahorrado en el último año es una práctica extendida en todos los niveles del índice, del 82,4 % entre quienes no aciertan ninguna pregunta al 92,8 % entre quienes aciertan las tres. La tenencia de productos, por el contrario, crece con mucha más intensidad. Disponer de al menos uno pasa del 15,7 % al 48,4 % entre los extremos del índice, las criptomonedas multiplican su presencia del 3,4 % al 22,6 % y el fondo de inversión, del 2,3 % al 16,5 %. Estos resultados apuntan en la dirección esperada por las hipótesis, pero no permiten atribuir las diferencias observadas exclusivamente al conocimiento financiero. Para ello es necesario controlar otros factores relevantes.

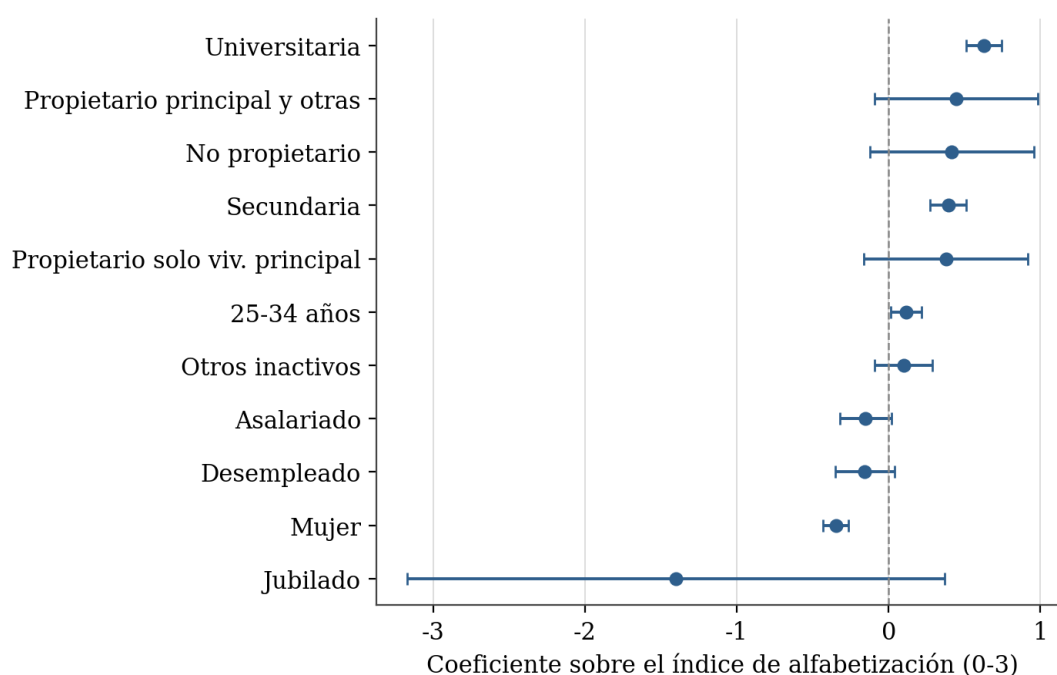
Tabla 3. Conducta financiera de los jóvenes según el número de aciertos (%)

Conducta	0 aciertos	1 acierto	2 aciertos	3 aciertos
Ha ahorrado en el último año	82,4	87,6	88,6	92,8
Tiene algún producto	15,7	23	32,1	48,4
Cuenta de ahorro	2,7	2,8	6,8	7,5
Plan de pensiones	1,9	2,4	6,5	11,6
Fondo de inversión	2,3	5	9,4	16,5
Acciones	0	0,3	0,8	1,2
Renta fija	7,7	10,1	14,4	18,3
Criptomonedas	3,4	8,8	11,6	22,6

Fuente: elaboración propia a partir de la Encuesta de Competencias Financieras 2021 (Banco de España y CNMV)

4.2. Determinantes del conocimiento financiero de los jóvenes

Las comparaciones de este apartado se entienden en igualdad de condiciones, es decir, entre personas que solo difieren en la característica examinada. La educación encabeza las diferencias, ya que, frente a quienes tienen estudios primarios, los titulados universitarios aciertan de media 0,63 preguntas más, y quienes alcanzan secundaria, 0,39.

Gráfica 4. Determinantes del conocimiento financiero (coeficientes del modelo de regresión lineal con intervalos de confianza al 95 %)

Fuente: elaboración propia a partir de la Encuesta de Competencias Financieras 2021 (Banco de España y CNMV)

La brecha de género es la segunda en magnitud. Las mujeres obtienen 0,35 aciertos menos que los hombres, una distancia amplia en una escala de cero a tres. Por su parte, el grupo de 25 a 34 años supera al de 18 a 24 en 0,11 puntos, una diferencia pequeña pero estadísticamente significativa.

El resto de variables aporta poco. Ninguna categoría laboral se distingue de la de referencia, los autónomos, y el coeficiente de los jubilados, un grupo casi inexistente a estas edades, resulta demasiado impreciso para interpretarlo. Las categorías de riqueza inmobiliaria tampoco alcanzan significación estadística individual.

El modelo explica en torno al 10 % de la variabilidad del índice ($R^2 = 0,096$), una capacidad modesta que indica que buena parte de las diferencias de conocimiento responde a factores no recogidos en la encuesta, como la socialización familiar o el interés personal por las finanzas.

4.3. Alfabetización y tenencia de productos de ahorro e inversión

Los modelos logit permiten evaluar las hipótesis incorporando las variables de control. La Tabla 4 presenta los *odds ratios* asociados al índice de alfabetización financiera para cada una de las conductas analizadas.

Tabla 4. Efecto del índice de alfabetización sobre la conducta financiera

Conducta	<i>Odds ratio</i>	IC 95 %	Significación
Haber ahorrado	1,16	0,99 – 1,36	Marginal ($p = 0,076$)
Algún producto	1,55	1,37 – 1,74	Sí ($p < 0,001$)
Cuenta de ahorro	1,44	1,13 – 1,85	Sí ($p = 0,003$)
Plan de pensiones	1,86	1,44 – 2,39	Sí ($p < 0,001$)
Fondo de inversión	1,63	1,33 – 2,00	Sí ($p < 0,001$)
Acciones	1,84	0,89 – 3,82	No ($p = 0,099$)
Renta fija	1,35	1,15 – 1,58	Sí ($p < 0,001$)
Criptomonedas	1,63	1,37 – 1,94	Sí ($p < 0,001$)

Fuente: elaboración propia a partir de la Encuesta de Competencias Financieras 2021 (Banco de España y CNMV)

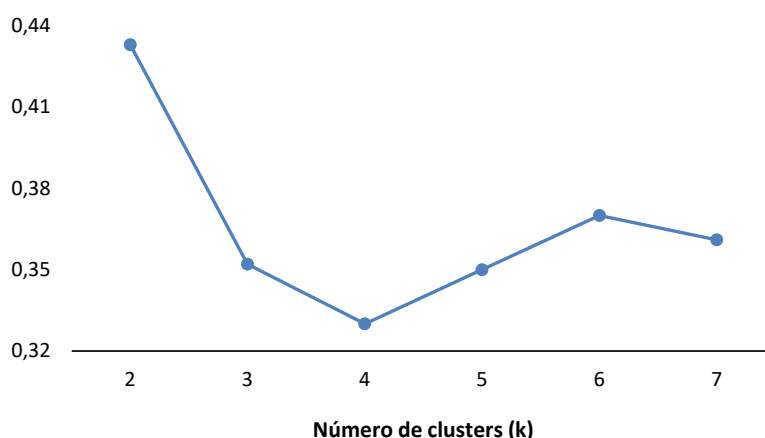
Los resultados respaldan la primera hipótesis, ya que la relación entre conocimiento y tenencia de productos es fuerte y consistente. Cada acierto adicional se asocia con un incremento del 55 % en las probabilidades relativas (odds) de disponer de al menos un producto financiero. Este patrón se repite en prácticamente todos los instrumentos, con el plan de pensiones a la cabeza (1,86), seguido de los fondos de inversión y las criptomonedas (ambos 1,63), las cuentas de ahorro (1,44) y la renta fija en último lugar (1,35), todos ellos significativos. Las acciones constituyen la única excepción. Su coeficiente es elevado (1,84) pero no alcanza significación estadística, dado que las posee menos del 1 % de los jóvenes.

El comportamiento del ahorro y el de la inversión, sin embargo, difieren de forma notable, confirmando así la segunda hipótesis. La asociación con el ahorro es débil, con un odds ratio de apenas 1,16 y una significación estadística limitada ($p = 0,076$). La tenencia de instrumentos, por el contrario, se mantiene clara y robusta en casi todos los casos. El conocimiento financiero influye, por tanto, poco en la decisión de guardar dinero, pero mucho en la de elegir los productos en los que invertirlo.

4.4. Perfiles de jóvenes según alfabetización y comportamiento

La segmentación mediante *k-means* reduce la cohorte a tres perfiles. Aunque el coeficiente de silueta favorecía una partición en dos grupos y, en segundo lugar, una de seis, ambas presentaban inconvenientes para los objetivos del estudio. La primera distinguía esencialmente entre quienes ahorran y quienes no, mientras que la segunda generaba grupos pequeños y difíciles de interpretar. Por ello, se optó por tres clústeres, una solución que ofrece perfiles más útiles desde el punto de vista económico. Conviene señalar, además, que el análisis combina una variable binaria (hábito de ahorro) con otras continuas, por lo que los resultados deben interpretarse con carácter exploratorio y descriptivo.

Gráfica 5. Evaluación del coeficiente de silueta en función del número de clúster (k)

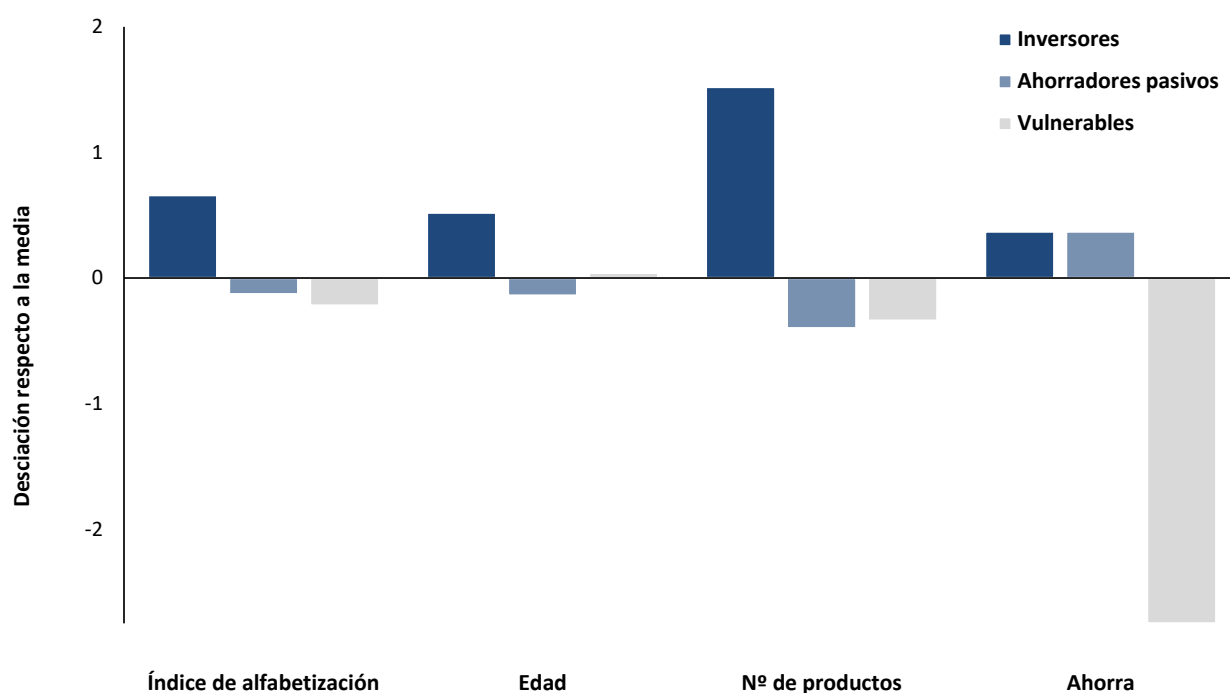


Fuente: elaboración propia a partir de la Encuesta de Competencias Financieras 2021 (Banco de España y CNMV)

El grupo mayoritario, dos de cada tres jóvenes, lo forman los ahorradores pasivos. Ahorran de forma habitual, pero tienen pocos productos financieros y un conocimiento medio. Los inversores, algo más de uno de cada cinco, concentran el mayor nivel de alfabetización y la edad media más alta, y son los únicos que combinan varios instrumentos. El tercer grupo, los vulnerables, agrupa al 12 % restante. No ahorran, carecen de productos y su conocimiento se parece mucho al de los pasivos. Es el perfil más expuesto a cualquier imprevisto económico.

Los perfiles matizan el papel del conocimiento. Lo que define a los inversores es su alfabetización, mientras que, entre pasivos y vulnerables, la diferencia está en el hábito de ahorro. Por esta razón, cada grupo requiere una respuesta diferente. Al bloque mayoritario le falta dar el paso hacia los instrumentos, en tanto que los vulnerables afrontan una carencia previa, más ligada a los recursos que a la formación.

Gráfica 6. Perfiles en puntuaciones estandarizadas (z-scores)



Fuente: elaboración propia a partir de la Encuesta de Competencias Financieras 2021 (Banco de España y CNMV)

4.5. Análisis de robustez

Los hallazgos anteriores se someten a tres comprobaciones de estabilidad frente a decisiones metodológicas alternativas.

- **Réplica ponderada:** al repetir los cálculos con el peso muestral, los coeficientes centrales apenas se mueven. La brecha de género queda en $-0,32$ y la ventaja universitaria en $+0,61$, y los *odds ratios* se mantienen entre el 1,15 de haber ahorrado y el 2,02 del plan de pensiones. En el plano descriptivo, el porcentaje de jóvenes con alfabetización plena baja del 18,2 % al 17,2 %, una cifra próxima al 19 % que publica el Banco de España (2023), aunque la comparación es solo orientativa, porque el dato oficial abarca a toda la población de 18 a 79 años.
- **Tratamiento alternativo del «no sabe»:** casi uno de cada cuatro jóvenes, el 23,5 %, responde «no sabe» a alguna de las tres preguntas. Al distinguir esas respuestas de las incorrectas en lugar de computarlas como fallo, la brecha de género se reduce en torno a un 31 %, de $-0,35$ a $-0,24$ puntos, aunque sigue siendo claramente significativa.
- **Corte de edad alternativo:** sobre las 1.322 observaciones de la definición de juventud restringida a los 18-29 años, la brecha de género, de $-0,37$, y la ventaja

universitaria, de +0,60, permanecen estables, y el *odds ratio* del índice para tener algún producto es idéntico al de la muestra principal, 1,55. Las conclusiones no dependen, por tanto, del corte elegido.

Las tres comprobaciones confirman que los resultados centrales del capítulo no son un artefacto de las decisiones tomadas al construir la muestra o el indicador.

5. DISCUSIÓN

5.1. Interpretación de los resultados

Los resultados confirman que los jóvenes españoles parten de un nivel algo inferior al del conjunto de la población. El panorama general encaja con lo que ya señalaba el marco teórico; la educación se confirma como el factor más determinante y la brecha de género reaparece con fuerza (Lusardi y Mitchell, 2014).

Ese rezago inicial admite una lectura menos pesimista de lo que aparenta. El retraso no nace en la juventud, sino que se arrastra desde la etapa escolar, donde el déficit ya resulta visible. Sin embargo, la distancia con el resto de la población se estrecha a medida que avanza la edad, lo que sugiere que el contacto con las primeras decisiones económicas aporta un aprendizaje que la enseñanza formal no llega a cubrir. El conocimiento financiero parece construirse, en buena medida, con la práctica.

Esa brecha de género, sin embargo, merece una lectura prudente. El análisis muestra que se sostiene incluso entre personas de igual formación y situación económica, lo que apunta a que existe una base real. Al mismo tiempo, parte de esa distancia parece venir de la forma de responder y no de un conocimiento realmente distinto. Ante una pregunta dudosa, las mujeres tienden a escoger la opción «no sabe» con más frecuencia, mientras que los hombres se inclinan por arriesgar una respuesta. Como el índice solo premia los aciertos, esa diferencia de actitud ensancha la distancia medida más de lo que justifican los conocimientos reales (Hospido et al., 2024). Interpretar la brecha como una pura diferencia de conocimiento sería, por tanto, exagerar su alcance.

No obstante, el hallazgo más importante es otro. El conocimiento financiero apenas guarda relación con la decisión de ahorrar, pero sí con la de dar un destino a ese dinero. La alfabetización no parece impulsar por sí sola el ahorro, actúa más bien como una puerta de entrada a los productos de inversión. Esta lectura encaja con la idea de que la falta de formación dificulta el acceso a los mercados financieros (van Rooij et al., 2011). En otras palabras, ahorrar depende sobre todo de tener capacidad económica, mientras que invertir exige, además, comprender.

Dentro de este marco, el caso de las criptomonedas resulta especialmente interesante. A menudo se perciben como un activo puramente especulativo, lo que podría sugerir un tenedor poco informado. Los datos, sin embargo, apuntan en sentido contrario, puesto que su tenencia se asocia a un mayor conocimiento. Ahora bien, ello no significa que invertir en criptoactivos sea una buena decisión, simplemente indica que quienes interactúan con estos activos suelen desenvolverse mejor en el ámbito financiero.

Los perfiles obtenidos ayudan a ordenar todo lo anterior. El conocimiento destaca claramente al grupo que invierte, pero no explica la diferencia entre quienes ahorran sin invertir y quienes ni siquiera lo hacen. La existencia de un grupo que ni ahorra ni invierte recuerda que las decisiones financieras de los jóvenes dependen tanto de lo que saben como de los medios con los que cuentan (Martínez Santos y Galindo, 2025).

5.2. Implicaciones

De este diagnóstico se desprende una orientación clara para la educación financiera. El esfuerzo no debería centrarse tanto en animar a ahorrar, algo que la mayoría de los jóvenes ya hace, sino en convertir ese ahorro en una decisión de inversión meditada. Una formación dirigida a ese momento concreto, centrada en los productos y en la relación entre riesgo y rentabilidad, resulta más prometedora que los mensajes genéricos (Kaiser et al., 2022).

La brecha de género reclama también una atención específica. Buena parte de esa distancia tiene que ver con la seguridad de las mujeres jóvenes a la hora de manejar sus propias finanzas, más que con su capacidad real. Esto se debe a que en torno a un tercio de la brecha de conocimiento se ha atribuido a la menor confianza de las mujeres y no a un saber inferior (Bucher-Koenen et al., 2021). Por eso conviene dirigir hacia ellas acciones concretas, desde una formación que fomente la confianza en la toma de decisiones hasta una mayor presencia de referentes femeninos en el ámbito financiero. El objetivo no es solo transmitir conocimientos, sino animar a aplicarlos.

El caso de los criptoactivos sugiere, por último, un enfoque distinto. Dado que quienes los poseen suelen estar bien informados, un mensaje basado en la falta de conocimiento resultaría poco eficaz. Resulta más útil concentrarse en los riesgos propios de estos activos y en la situación de quienes invierten en ellos una buena parte de su patrimonio.

5.3. Limitaciones

Conviene leer los resultados anteriores teniendo presentes varias limitaciones.

- **Naturaleza de los datos:** la encuesta ofrece una fotografía de un único momento, de modo que las asociaciones halladas no permiten afirmar causalidad. Podría incluso ocurrir lo contrario, que la experiencia de invertir aumente el conocimiento. Además, puede haber factores no observados que influyan simultáneamente en ambas conductas.
- **Medida de la alfabetización financiera:** un indicador de tres preguntas resume una competencia mucho más amplia y, como se ha visto, es sensible a la forma de tratar las respuestas dudosas. Tampoco capta la habilidad práctica, que es la más próxima a las decisiones reales.
- **Aproximación a los recursos:** al quedar fuera la renta del hogar, la posición económica se mide de forma indirecta. Parte de lo atribuido al conocimiento podría reflejar diferencias económicas no recogidas del todo.
- **Medida de la conducta:** la tenencia se registra como presencia o ausencia, sin reflejar importes ni acierto de las decisiones. Tener un producto no significa usarlo bien, y la escasez de algún activo impide estimaciones precisas.
- **Carácter exploratorio de la segmentación:** los perfiles dependen de las variables elegidas y del número de grupos fijado, de manera que describen de forma ordenada la diversidad interna, pero no constituyen categorías cerradas.

Ninguna de estas limitaciones desmonta el resultado principal, que se mantiene firme en las distintas comprobaciones, pero todas invitan a interpretarlo con la debida prudencia.

6. CONCLUSIONES

6.1. Síntesis de los resultados

Este trabajo se propuso analizar la relación entre la alfabetización financiera de los jóvenes españoles y sus decisiones de ahorro e inversión. Los cuatro objetivos iniciales permiten recapitular lo aprendido.

El primero buscaba situar el nivel de conocimiento de los jóvenes. Se ha comprobado que la mayoría se concentra en una posición intermedia, algo por debajo del conjunto de la población, con un rezago que asoma ya en la etapa escolar pero que muestra signos de mejora. Esa posición, ni alta ni alarmante, sugiere que el problema no es tanto una carencia generalizada como una preparación incompleta frente a decisiones cada vez más complejas.

El segundo perseguía identificar los factores que explican las diferencias internas. La formación académica resultó ser el factor más decisivo, acompañado de una brecha de género que se mantiene al comparar individuos equiparables. La edad influye menos, mientras que la situación laboral y los recursos del hogar apenas añaden capacidad explicativa.

El tercero, el más importante, indagaba si el conocimiento se traduce en tenencia de productos. La respuesta separa con claridad dos decisiones, la de ahorrar, donde el conocimiento pesa poco, y la de invertir, donde su influencia se vuelve determinante.

El cuarto proponía agrupar a los jóvenes en perfiles. Surgieron tres; un amplio grupo que ahorra sin dar el salto a la inversión, una minoría inversora más formada y un grupo en situación frágil que ni ahorra ni participa. Lo relevante es que la línea entre estos dos últimos no la marca el conocimiento, sino la posibilidad de ahorrar. La segmentación confirma así, desde otro ángulo, el mismo mensaje, que la formación distingue a quien invierte, pero son los recursos los que separan a quien puede ahorrar de quien no.

6.2. Respuesta a las hipótesis

El conjunto de resultados permite resolver las dos hipótesis planteadas en el marco teórico.

La primera esperaba una asociación positiva entre alfabetización y tenencia de productos, una vez tenidas en cuenta las características personales. Los datos la respaldan, pues esa asociación se mostró sólida y constante a lo largo de las comprobaciones realizadas. El vínculo se mantuvo al emplear el peso muestral, al tratar de otro modo las respuestas dudosas y al restringir la edad de la muestra, lo que refuerza la confianza en este resultado. Un mayor conocimiento financiero se asocia, por tanto, con una mayor presencia de productos de ahorro e inversión entre los jóvenes españoles.

La segunda anticipaba que el conocimiento importaría más para elegir instrumentos que para ahorrar. También queda confirmada, y en ella reside la principal enseñanza del trabajo; la alfabetización financiera no determina si un joven ahorra, sino cómo decide emplear lo ahorrado. La distancia entre una y otra decisión resultó clara, pues el conocimiento apenas se relaciona con el hecho de ahorrar pero sí con la tenencia de casi todos los productos de inversión. Esta diferencia matiza la idea, extendida, de que más formación financiera conduce sin más a un mayor ahorro. Lo que el conocimiento parece favorecer no es ahorrar, sino dar a ese ahorro un destino más elaborado.

6.3. Aportaciones y líneas futuras

La contribución de este estudio consiste en haber separado con nitidez dos decisiones que suelen tratarse juntas, la de ahorrar y la de invertir. Distinguir las revela que el conocimiento actúa sobre la segunda mucho más que sobre la primera, un matiz que se pierde cuando ambas se examinan en bloque. A ello se añade el uso de la edición más reciente de la encuesta, que ofrece una imagen actualizada de un grupo en plena transformación. El trabajo aporta, además, una mirada centrada de forma específica en la población joven, un colectivo que los análisis generales tienden a diluir en el conjunto de la población adulta.

El análisis también señala posibles líneas futuras. Disponer de datos que sigan a las mismas personas a lo largo del tiempo permitiría acercarse a la causalidad, hoy fuera de alcance, y distinguir si es el conocimiento el que lleva a invertir o la propia experiencia de invertir la que genera conocimiento. Incorporar información más detallada sobre la conducta, que recoja cuánto se invierte y no solo si se hace, mejoraría el análisis. Resultaría valioso, asimismo, profundizar en el papel de la confianza, que este trabajo solo aborda de forma indirecta, para comprender mejor por qué el conocimiento se traduce en conducta en unos casos

y no en otros. Por último, convendría observar a esta generación en futuras ediciones de la encuesta. Así podría comprobarse si acorta distancias con el resto de la población y si el vínculo entre conocimiento e inversión se mantiene en un entorno económico cambiante.

Bibliografía

- Allgood, S. y Walstad, W. B. (2016). The effects of perceived and actual financial literacy on financial behaviors. *Economic Inquiry*, 54(1), 675-697.
- Banco de España (2023). *Encuesta de Competencias Financieras 2021*. Madrid: Banco de España.
- Bover, O., Hospido, L. y Villanueva, E. (2018). The impact of high school financial education on financial knowledge and choices: Evidence from a randomized trial in Spain. Documentos de Trabajo n.º 1801, Banco de España.
- Bucher-Koenen, T., Alessie, R., Lusardi, A. y van Rooij, M. (2021). Fearless woman: Financial literacy and stock market participation. NBER Working Paper 28723.
- CNMV y Banco de España (2022). *Plan de Educación Financiera 2022-2025*. Madrid.
- Consejo de la Juventud de España (2026). *Observatorio de Emancipación*. Madrid: CJE.
- de Bassa Scheresberg, C. (2013). Financial literacy and financial behavior among young adults: Evidence and implications. *Numeracy*, 6(2).
- Fernandes, D., Lynch, J. G. y Netemeyer, R. G. (2014). Financial literacy, financial education, and downstream financial behaviors. *Management Science*, 60(8), 1861-1883.
- Funcas (2024). *Educación financiera de los adolescentes españoles*. Notas del Observatorio de la Digitalización Financiera. Madrid: Funcas.
- Guirola, L. M. e Isla Martínez, C. (2025). La tenencia de criptoactivos entre los hogares españoles. *Boletín Económico - Banco de España*, 2025/T3, artículo 06. <https://doi.org/10.53479/40588>.
- Hospido, L., Iriberry, N. y Machelett, M. (2024). Gender gaps in financial literacy: A multi-arm RCT to break the response bias in surveys. Documentos de Trabajo n.º 2401, Banco de España.
- Hsu, J. W. (2011). Aging and strategic learning: The impact of spousal incentives on financial literacy. Finance and Economics Discussion Series 2011-53, Board of Governors of the Federal Reserve System.
- Huston, S. J. (2010). Measuring financial literacy. *Journal of Consumer Affairs*, 44(2), 296-316.
- Ispuerto Maté, A., Martínez García, I. y Ruiz Suárez, G. (2021). Educación financiera y decisiones de ahorro e inversión: un análisis de la Encuesta de Competencias Financieras (ECF). Documentos de Trabajo n.º 75, CNMV.

- Kaiser, T., Lusardi, A., Menkhoff, L. y Urban, C. (2022). Financial education affects financial knowledge and downstream behaviors. *Journal of Financial Economics*, 145(2), 255-272.
- Lusardi, A. (2019). Financial literacy and the need for financial education: Evidence and implications. *Swiss Journal of Economics and Statistics*, 155(1), artículo 1.
- Lusardi, A. y Mitchell, O. S. (2011). Financial literacy around the world: An overview. *Journal of Pension Economics and Finance*, 10(4), 497-508.
- Lusardi, A. y Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature*, 52(1), 5-44.
- Lusardi, A., Mitchell, O. S. y Curto, V. (2010). Financial literacy among the young. *Journal of Consumer Affairs*, 44(2), 358-380.
- Mancebón Torrubia, M. J., Ximénez-de-Embún, D. P. y Villar-Aldonza, A. (2020). Habilidades financieras y hábitos financieros saludables: un análisis a partir de la Encuesta de Competencias Financieras. *Cuadernos de Información Económica*, 275, 55-69.
- Martínez Santos, J. y Galindo, J. (2025). *Radiografía de la evolución del patrimonio de los hogares españoles: la dificultad del ahorro para las nuevas generaciones*. Observatorio del Ahorro Familiar.
- OCDE (2024). *PISA 2022 Results (Volume IV): How Financially Smart Are Students?* París: OECD Publishing.
- OCDE/INFE (2012). *High-level Principles on National Strategies for Financial Education*. París: OECD Publishing.
- van Rooij, M., Lusardi, A. y Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial Economics*, 101(2), 449-472.

Anexo 1

Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado

ADVERTENCIA: Desde la Universidad consideramos que ChatGPT u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea.

Por la presente, yo, Lucas Esnal, estudiante de E2 + Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado " DISEÑO DEL PLAN DE NEGOCIO DE UNA RED SOCIAL ORIENTADA A LA INVERSIÓN Y LA EDUCACIÓN FINANCIERA: INVESNET", 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. **Crítico:** Para encontrar contra-argumentos a una tesis específica que pretendo defender.
3. **Referencias:** Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
4. **Metodólogo:** Para descubrir métodos aplicables a problemas específicos de investigación.
5. **Interpretador de código:** Para realizar análisis de datos preliminares.
6. **Corrector de estilo literario y de lenguaje:** Para mejorar la calidad lingüística y estilística del texto.
7. **Sintetizador y divulgador de libros complicados:** Para resumir y comprender literatura compleja.
8. **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: Junio 2026

Firma: Lucas Esnal Olarreaga

Análisis empírico

Cuaderno de análisis de la relación entre alfabetización financiera y decisiones de ahorro e inversión en los jóvenes españoles de 18 a 34 años, a partir de la Encuesta de Competencias Financieras de 2021 del Banco de España.

Carga y preparación de los datos

Parto del fichero de microdatos en formato Stata y compruebo que las variables que voy a utilizar están presentes antes de empezar.

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

import statsmodels.api as sm
import statsmodels.formula.api as smf

from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
from sklearn.metrics import silhouette_score

# Estética para que las figuras encajen bien en el TFG
sns.set_theme(style="whitegrid", context="notebook")
plt.rcParams.update({
    "figure.figsize": (8, 5),
    "figure.dpi": 110,
    "font.size": 11,
    "savefig.bbox": "tight",
})

# Semilla para que los resultados con aleatoriedad (k-means) sean reproducibles
RANDOM_STATE = 42
```

```
In [2]: RUTA_ECF_2021 = "ecf_2021.dta"

df = pd.read_stata(RUTA_ECF_2021, convert_categoricals=False)

print(f"Filas: {df.shape[0]:,}")
print(f"Columnas: {df.shape[1]:,}")
df.head()
```

Filas: 7,764

Columnas: 429

```
Out[2]:
```

	a01	a02	a0000	a0400	a04	a0800	a0100	a0320	a1030	a1040	...	a
0	2022	6	0	1982	40	-98	1	0	2	2	...	a

1	2022	3	0	1967	54	-98	1	0	1	-98	...
2	2022	2	1	1967	54	-98	1	0	6	-98	...
3	2022	5	1	1974	48	-98	1	0	2	2	...
4	2022	6	1	1972	49	-98	1	0	1	-98	...

5 rows × 429 columns

```
In [3]: vars_esperadas = [
# Demográficas
"age", "a1100", "a1500", "a2000",
# Big Three
"e0600", "e0900", "e1003",
# Tenencia de productos
"b0301", "b0302", "b0303", "b0304", "b0305", "b0306", "b0308", "b0312",
# Comportamiento de ahorro
"b1000j",
# Fragilidad financiera
"j0200",
# Vivienda
"i0100", "i1000",
]

faltantes = [v for v in vars_esperadas if v not in df.columns]
presentes = [v for v in vars_esperadas if v in df.columns]

print(f"Variables presentes: {len(presentes)}/{len(vars_esperadas)}")
if faltantes:
    print("Faltan:", faltantes)
```

Variables presentes: 19/19

Recodificación de variables

Recodifico las variables sociodemográficas siguiendo los criterios del Banco de España, de modo que mis resultados sean comparables con los oficiales. Los códigos -97, -98, -99 y -5 corresponden a "no sabe", "no contesta" y preguntas no aplicables.

```
In [4]: CODIGOS_NSNC = [-97, -98, -99, -5]

def es_nsnc(serie):
    """Devuelve una máscara booleana donde la respuesta es NS/NC/filtrada."""
    return serie.isin(CODIGOS_NSNC)
```

He identificado el sexo a partir de la variable `a0000`. Como no traía etiquetas de texto, confirmé la codificación comprobando que el grupo con mayor porcentaje de aciertos coincide con los hombres, tal como describe la literatura.

```
In [5]: # a0000: 1 = hombre, 0 = mujer (confirmado por % de aciertos vs. cifras del BdE)
df["sexo"] = df["a0000"].map({1: "Hombre", 0: "Mujer"})
```

```
In [6]: # Tramos de edad usados por el BdE
df["age_bracket"] = pd.cut(
    df["age"],
    bins=[0, 34, 44, 54, 64, 79],
    labels=["18-34", "35-44", "45-54", "55-64", "65-79"],
)
```

```
In [7]: def recodifica_educacion(x):
    if x in [1, 2, 3]:
        return 1 # primaria
    if x in [4, 5]:
        return 2 # secundaria
    if x in [6, 7, 8, 9]:
        return 3 # universitaria
    return np.nan

df["edu_level"] = df["a1100"].apply(recodifica_educacion)
df["edu_label"] = df["edu_level"].map({1: "Primaria", 2: "Secundaria", 3: "Universi
```

```
In [8]: df["labor_status"] = np.nan
df.loc[df["a1500"] == 1, "labor_status"] = 1 # autónomo
df.loc[(df["a2000"] == 1) | (df["a1500"] == 2), "labor_status"] = 2 # asalariado
df.loc[df["a1500"] == 5, "labor_status"] = 3 # desempleado
df.loc[df["a1500"] == 6, "labor_status"] = 4 # jubilado
df.loc[df["a1500"].isin([4, 7, 8, 9, 10]), "labor_status"] = 5 # otros inactivos

df["labor_label"] = df["labor_status"].map({
    1: "Autónomo", 2: "Asalariado", 3: "Desempleado",
    4: "Jubilado", 5: "Otros inactivos"
})
```

```
In [9]: df["riqueza_in"] = 0
df.loc[df["i0100"].isin([3, 4]), "riqueza_in"] = 1 # no propietario
df.loc[(df["i0100"].isin([1, 2])) & (df["i1000"] == 0), "riqueza_in"] = 2 # solo p
df.loc[(df["i0100"].isin([1, 2])) & (df["i1000"] == 1), "riqueza_in"] = 3 # princi
```

Índice de alfabetización financiera

Construyo el indicador a partir de las tres preguntas de Lusardi y Mitchell sobre inflación, interés compuesto y diversificación del riesgo. Distingo las respuestas correctas, las incorrectas y las de "no sabe", y sumo los aciertos para obtener un índice de 0 a 3.

```
In [10]: # Pregunta 1: Inflación (correcta si e0600 ∈ {3, 4})
df["correct_e0600"] = (df["e0600"].isin([3, 4])).astype(int)
df["dk_e0600"] = (df["e0600"] == -97).astype(int)
df["da_e0600"] = (df["e0600"].isin([-99, -5])).astype(int)
```

```

df["incorrect_e0600"] = 1 - df["correct_e0600"] - df["dk_e0600"] - df["da_e0600"]

# Pregunta 2: Interés compuesto (correcta si e0900 == 1)
df["correct_e0900"] = (df["e0900"] == 1).astype(int)
df["dk_e0900"] = (df["e0900"] == -97).astype(int)
df["da_e0900"] = (df["e0900"].isin([-99, -5])).astype(int)
df["incorrect_e0900"] = 1 - df["correct_e0900"] - df["dk_e0900"] - df["da_e0900"]

# Pregunta 3: Diversificación del riesgo (correcta si e1003 == 1)
df["correct_e1003"] = (df["e1003"] == 1).astype(int)
df["dk_e1003"] = (df["e1003"] == -97).astype(int)
df["da_e1003"] = (df["e1003"].isin([-99, -5])).astype(int)
df["incorrect_e1003"] = 1 - df["correct_e1003"] - df["dk_e1003"] - df["da_e1003"]

```

```

In [11]: df["indice_alfab"] = df[["correct_e0600", "correct_e0900", "correct_e1003"]].sum(axis=1)
df["alfab_plena"] = (df["indice_alfab"] == 3).astype(int)
df["indice_alfab_n"] = df["indice_alfab"] / 3

# Distribución general (todas las edades) como referencia
print("Distribución del índice de alfabetización (toda la muestra):")
print(df["indice_alfab"].value_counts(normalize=True).sort_index().mul(100).round(1))

```

Distribución del índice de alfabetización (toda la muestra):

indice_alfab

0 12.3

1 31.9

2 35.0

3 20.7

Name: proportion, dtype: float64

Productos financieros, ahorro y fragilidad

Defino la tenencia de cada producto y construyo dos indicadores agregados: disponer de algún vehículo de ahorro o inversión y tener alguna deuda. Añado la variable de ahorro en los últimos doce meses y un indicador de fragilidad financiera del hogar.

```

In [12]: # Tenencia individual de cada producto
df["t_hipoteca"] = (df["b0301"] == 1).astype(int)
df["t_cuenta_ahorro"] = (df["b0302"] == 1).astype(int)
df["t_plan_pensiones"] = (df["b0303"] == 1).astype(int)
df["t_fondo_inv"] = (df["b0304"] == 1).astype(int)
df["t_acciones"] = (df["b0305"] == 1).astype(int)
df["t_prest_personal"] = (df["b0306"] == 1).astype(int)
df["t_renta_fija"] = (df["b0308"] == 1).astype(int)
df["t_cripto"] = (df["b0312"] == 1).astype(int)

# Variables sintéticas siguiendo la sintaxis del BdE
df["t_alahoc"] = (
    (df["b0302"] == 1) | (df["b0303"] == 1) | (df["b0304"] == 1) |
    (df["b0305"] == 1) | (df["b0308"] == 1) | (df["b0312"] == 1)
).astype(int)

```

```
df["t_alhip"] = ((df["b0301"] == 1) | (df["b0306"] == 1)).astype(int)

# Conteo de productos de ahorro/inversión
df["n_prod_ahorro"] = df[[
    "t_cuenta_ahorro", "t_plan_pensiones", "t_fondo_inv",
    "t_acciones", "t_renta_fija", "t_cripto"
]].sum(axis=1)

# Ahorro en los últimos 12 meses
df["ahorra"] = np.where(df["b1000j"] == 0, 1, np.where(df["b1000j"] == 1, 0, np.nan))
```

```
In [13]: df["fragil"] = np.where(df["j0200"] == 1, 1, np.where(df["j0200"] == 0, 0, np.nan))
```

Delimitación de la cohorte joven

Restrinjo la muestra a los individuos de 18 a 34 años y los divido en dos subgrupos, de 18 a 24 y de 25 a 34, que corresponden a etapas distintas de la transición a la vida adulta.

```
In [14]: df["joven"] = (df["age"].between(18, 34)).astype(int)

df_joven = df[df["joven"] == 1].copy()

df_joven["subgrupo_edad"] = pd.cut(
    df_joven["age"],
    bins=[17, 24, 34],
    labels=["18-24", "25-34"]
)

print(f"Submuestra de jóvenes (18-34): {len(df_joven):,} observaciones")
print(f"Submuestra del resto de la población (35+): {(df['joven'] == 0).sum():,}")
print()
print("Distribución por subgrupo:")
print(df_joven["subgrupo_edad"].value_counts())
```

Submuestra de jóvenes (18-34): 1,899 observaciones

Submuestra del resto de la población (35+): 5,865

Distribución por subgrupo:

subgrupo_edad

25-34 1106

18-24 793

Name: count, dtype: int64

Análisis descriptivo

Empiezo por caracterizar a la cohorte y situar su nivel de conocimiento financiero frente al del resto de la población.

```
In [15]: perfil = pd.DataFrame({
    "Variable": ["Mujeres (%)", "Edad media", "Universitarios (%)",
```

```

        "Asalariados (%)", "Desempleados (%)",
        "Vive en vivienda en propiedad (%)"],
    "Valor": [
        round((df_joven["sexo"] == "Mujer").mean() * 100, 1),
        round(df_joven["age"].mean(), 1),
        round((df_joven["edu_level"] == 3).mean() * 100, 1),
        round((df_joven["labor_status"] == 2).mean() * 100, 1),
        round((df_joven["labor_status"] == 3).mean() * 100, 1),
        round((df_joven["riqueza_in"] >= 2).mean() * 100, 1),
    ]
})
perfil

```

Out[15]:

	Variable	Valor
0	Mujeres (%)	48.0
1	Edad media	26.1
2	Universitarios (%)	43.9
3	Asalariados (%)	50.8
4	Desempleados (%)	13.0
5	Vive en vivienda en propiedad (%)	64.4

A continuación comparo la distribución del índice entre los jóvenes y el resto de la población.

```

In [16]: fig, ax = plt.subplots(1, 2, figsize=(13, 4.5))

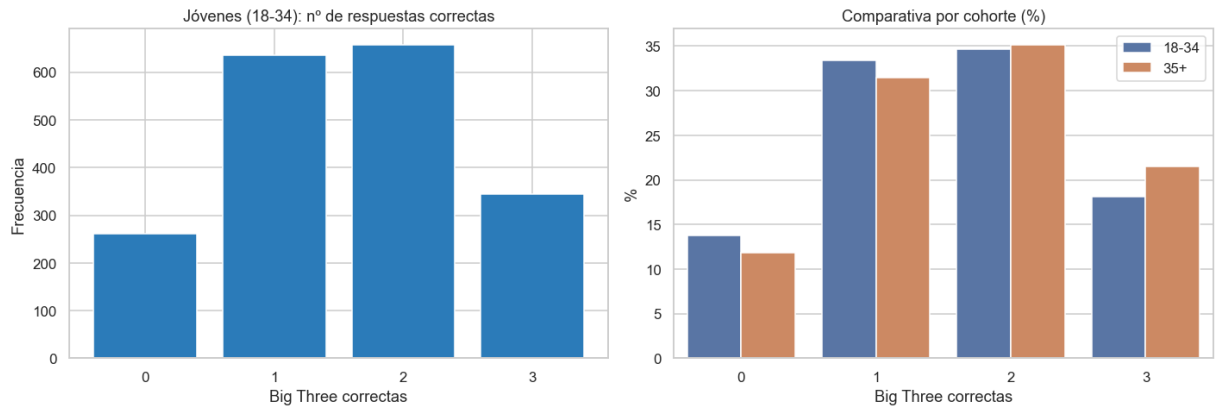
# Distribución absoluta en jóvenes
counts = df_joven["indice_alfab"].value_counts().sort_index()
ax[0].bar(counts.index, counts.values, color="#2b7bb9")
ax[0].set_title("Jóvenes (18-34): nº de respuestas correctas")
ax[0].set_xlabel("Big Three correctas")
ax[0].set_ylabel("Frecuencia")
ax[0].set_xticks([0, 1, 2, 3])

# Comparativa joven vs resto, en porcentaje
df["cohorta"] = np.where(df["joven"] == 1, "18-34", "35+")
prop = (df.groupby("cohorta")["indice_alfab"]
        .value_counts(normalize=True)
        .mul(100).rename("pct").reset_index())

sns.barplot(data=prop, x="indice_alfab", y="pct", hue="cohorta", ax=ax[1])
ax[1].set_title("Comparativa por cohorte (%)")
ax[1].set_xlabel("Big Three correctas")
ax[1].set_ylabel("%")
ax[1].legend(title="")

plt.tight_layout()
plt.show()

```



Cruzo después el nivel de alfabetización con la tenencia de cada producto, para observar si la asociación entre conocimiento y comportamiento aparece ya en los descriptivos.

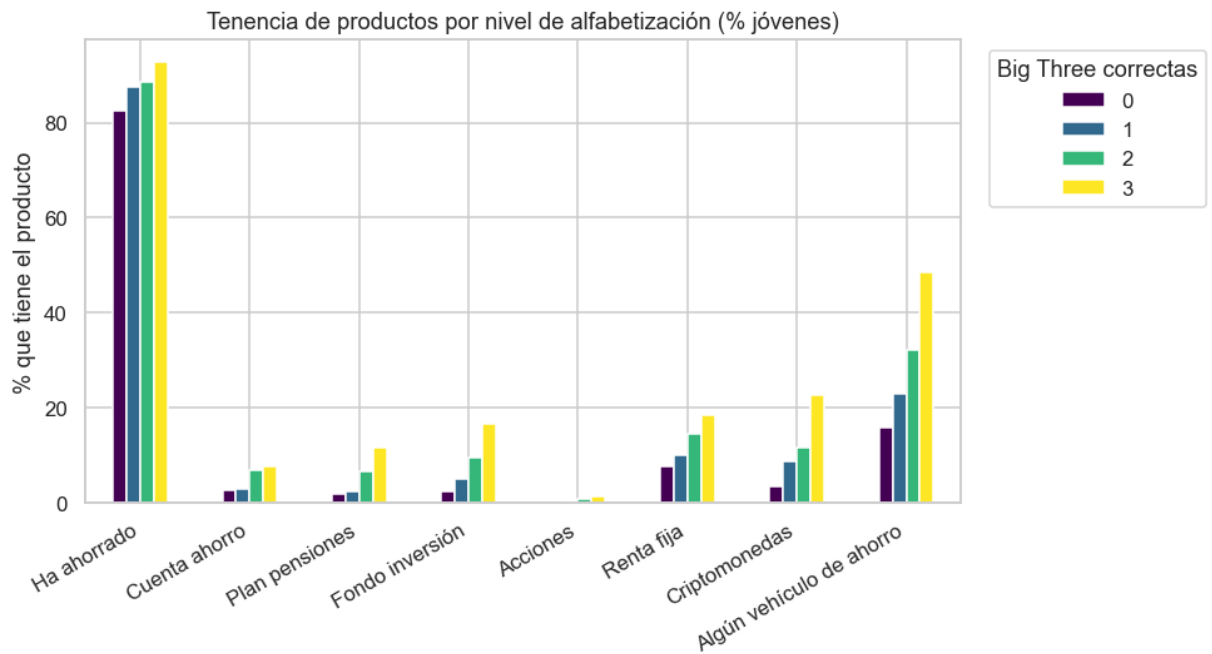
```
In [17]: productos = ["ahorra", "t_cuenta_ahorro", "t_plan_pensiones",
                    "t_fondo_inv", "t_acciones", "t_renta_fija", "t_cripto", "t_alahoc"]

tabla_cruzada = (df_joven.groupby("indice_alfab")[productos]
                .mean().mul(100).round(1))
tabla_cruzada.columns = ["Ha ahorrado", "Cuenta ahorro", "Plan pensiones",
                        "Fondo inversión", "Acciones", "Renta fija", "Criptomonedas",
                        "Algún vehículo de ahorro"]
tabla_cruzada
```

Out[17]:

	Ha ahorrado	Cuenta ahorro	Plan pensiones	Fondo inversión	Acciones	Renta fija	Cripto
indice_alfab							
0	82.4	2.7	1.9	2.3	0.0	7.7	
1	87.6	2.8	2.4	5.0	0.3	10.1	
2	88.6	6.8	6.5	9.4	0.8	14.4	
3	92.8	7.5	11.6	16.5	1.2	18.3	

```
In [18]: # Visualización compacta de la tabla anterior
fig, ax = plt.subplots(figsize=(9, 5))
tabla_cruzada.T.plot(kind="bar", ax=ax, colormap="viridis")
ax.set_title("Tenencia de productos por nivel de alfabetización (% jóvenes)")
ax.set_ylabel("% que tiene el producto")
ax.set_xlabel("")
ax.legend(title="Big Three correctas", bbox_to_anchor=(1.02, 1), loc="upper left")
plt.xticks(rotation=30, ha="right")
plt.tight_layout()
plt.show()
```



Determinantes de la alfabetización financiera

Estimo una regresión lineal del índice sobre las características sociodemográficas para identificar qué factores se asocian con un mayor nivel de conocimiento.

```
In [19]: # Convertimos las variables categóricas para que statsmodels las trate como dummies
for v in ["sexo", "edu_level", "labor_status", "riqueza_in", "subgrupo_edad"]:
    if v in df_joven.columns:
        df_joven[v] = df_joven[v].astype("category")

# Variables de control del modelo
controles = "C(sexo) + C(edu_level) + C(labor_status) + C(subgrupo_edad) + C(riqueza_in)"

formula = f"indice_alfab ~ {controles}"

modelo_lineal = smf.ols(formula, data=df_joven).fit()
print(modelo_lineal.summary())
```

OLS Regression Results

```
=====
Dep. Variable:      indice_alfab    R-squared:      0.096
Model:              OLS             Adj. R-squared: 0.091
Method:             Least Squares   F-statistic:    18.29
Date:               Wed, 17 Jun 2026 Prob (F-statistic): 4.14e-35
Time:               19:47:08        Log-Likelihood: -2480.2
No. Observations:   1899            AIC:            4984.
Df Residuals:       1887            BIC:            5051.
Df Model:           11
Covariance Type:    nonrobust
=====
```

```
=====
coef    std err        t    P>|t|    [0.025    0.975]
```

```
-----
-----
Intercept                0.9087    0.287    3.166    0.002    0.346
1.472
C(sexo)[T.Mujer]        -0.3478    0.042   -8.325    0.000   -0.430
-0.266
C(edu_level)[T.2]        0.3934    0.060    6.532    0.000    0.275
0.512
C(edu_level)[T.3]        0.6284    0.059   10.562    0.000    0.512
0.745
C(labor_status)[T.2.0]   -0.1522    0.086   -1.762    0.078   -0.322
0.017
C(labor_status)[T.3.0]   -0.1572    0.100   -1.573    0.116   -0.353
0.039
C(labor_status)[T.4.0]   -1.4006    0.901   -1.554    0.120   -3.168
0.367
C(labor_status)[T.5.0]    0.0980    0.097    1.010    0.313   -0.092
0.288
C(subgrupo_edad)[T.25-34] 0.1149    0.052    2.195    0.028    0.012
0.218
C(riqueza_in)[T.1]       0.4161    0.275    1.513    0.130   -0.123
0.955
C(riqueza_in)[T.2]       0.3771    0.274    1.375    0.169   -0.161
0.915
C(riqueza_in)[T.3]       0.4453    0.274    1.624    0.105   -0.093
0.983
=====
Omnibus:                125.139    Durbin-Watson:          1.955
Prob(Omnibus):           0.000    Jarque-Bera (JB):       46.862
Skew:                    -0.094    Prob(JB):               6.67e-11
Kurtosis:                2.254    Cond. No.               72.1
=====
```

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Represento los coeficientes con sus intervalos de confianza para visualizar su magnitud y su signo.

```
In [20]: def coef_plot(modelo, titulo="Coeficientes y IC 95%"):
    coefs = modelo.params.drop("Intercept", errors="ignore")
    ic = modelo.conf_int().drop("Intercept", errors="ignore")
    ic.columns = ["low", "high"]

    df_plot = pd.DataFrame({
        "var": coefs.index,
        "coef": coefs.values,
        "low": ic["low"].values,
        "high": ic["high"].values,
    }).sort_values("coef")

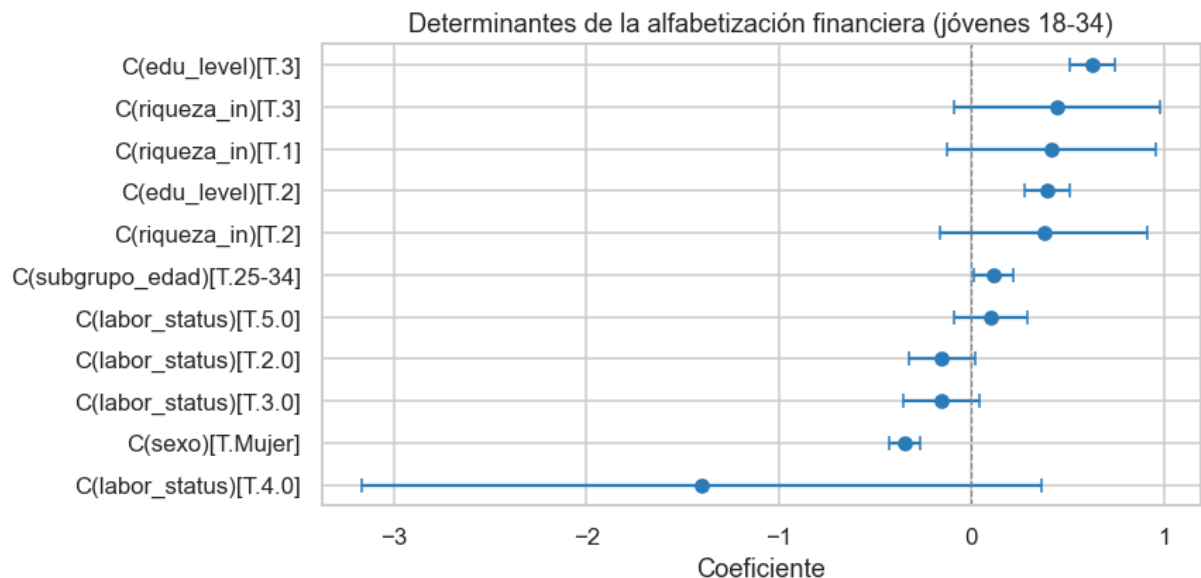
    fig, ax = plt.subplots(figsize=(8, max(4, 0.3 * len(df_plot))))
    ax.errorbar(df_plot["coef"], df_plot["var"],
                xerr=[df_plot["coef"] - df_plot["low"],
```

```

        df_plot["high"] - df_plot["coef"]],
        fmt="o", color="#2b7bb9", capsize=3)
ax.axvline(0, color="grey", lw=0.8, ls="--")
ax.set_title(titulo)
ax.set_xlabel("Coeficiente")
plt.tight_layout()
plt.show()

```

coef_plot(modelo_lineal, "Determinantes de la alfabetización financiera (jóvenes 18-34)



Alfabetización y decisiones de ahorro e inversión

Para las variables de comportamiento, que son binarias, utilizo regresión logística. Estimo un modelo por producto, manteniendo el índice de alfabetización como variable explicativa principal y las mismas variables de control.

```

In [21]: def logit_producto(df, dep_var, controles):
    """Ajusta un logit con la alfabetización como variable explicativa principal.
    Usamos el optimizador 'bfgs' con más iteraciones, más robusto para productos
    poco frecuentes (p. ej. acciones o renta fija, con muy pocos casos)."""
    formula = f"{dep_var} ~ indice_alfab + {controles}"
    return smf.logit(formula, data=df).fit(dis=False, method="bfgs", maxiter=500)

# t_alahoc (algún vehículo de ahorro) es la variable agregada más robusta.
# t_renta_fija y t_acciones tienen muy pocos casos: si no convergen o son
# residuales, se eliminan de las tablas en lugar de reportar un OR inestable.
vars_dep = ["ahorra", "t_alahoc", "t_cuenta_ahorro", "t_plan_pensiones",
            "t_fondo_inv", "t_acciones", "t_renta_fija", "t_cripto"]

resultados = {}

```

```

for v in vars_dep:
    if v in df_joven.columns:
        try:
            resultados[v] = logit_producto(df_joven, v, controles)
        except Exception as e:
            print(f"Error en {v}: {e}")

print(f"Modelos ajustados: {list(resultados.keys())}")

```

Modelos ajustados: ['ahorra', 't_alahoc', 't_cuenta_ahorro', 't_plan_pensiones', 't_fondo_inv', 't_acciones', 't_renta_fija', 't_cripto']

Reúno los resultados en una tabla de odds ratios y los represento en un diagrama que permite comparar el efecto de la alfabetización sobre cada producto.

```

In [22]: def tabla_odds_ratios(resultados, var_principal="indice_alfab"):
    filas = []
    nombres_legibles = {
        "ahorra": "Ha ahorrado (12m)",
        "t_alahoc": "Algún vehículo de ahorro",
        "t_cuenta_ahorro": "Cuenta de ahorro",
        "t_plan_pensiones": "Plan de pensiones",
        "t_fondo_inv": "Fondo de inversión",
        "t_acciones": "Acciones",
        "t_renta_fija": "Renta fija",
        "t_cripto": "Criptomonedas",
    }
    for v, mod in resultados.items():
        if var_principal in mod.params.index:
            coef = mod.params[var_principal]
            ic = mod.conf_int().loc[var_principal]
            filas.append({
                "Producto": nombres_legibles.get(v, v),
                "OR": round(np.exp(coef), 2),
                "IC 95% bajo": round(np.exp(ic[0]), 2),
                "IC 95% alto": round(np.exp(ic[1]), 2),
                "p-valor": round(mod.pvalues[var_principal], 4),
                "N": int(mod.nobs),
            })
    return pd.DataFrame(filas)

tabla_or = tabla_odds_ratios(resultados)
tabla_or

```

```

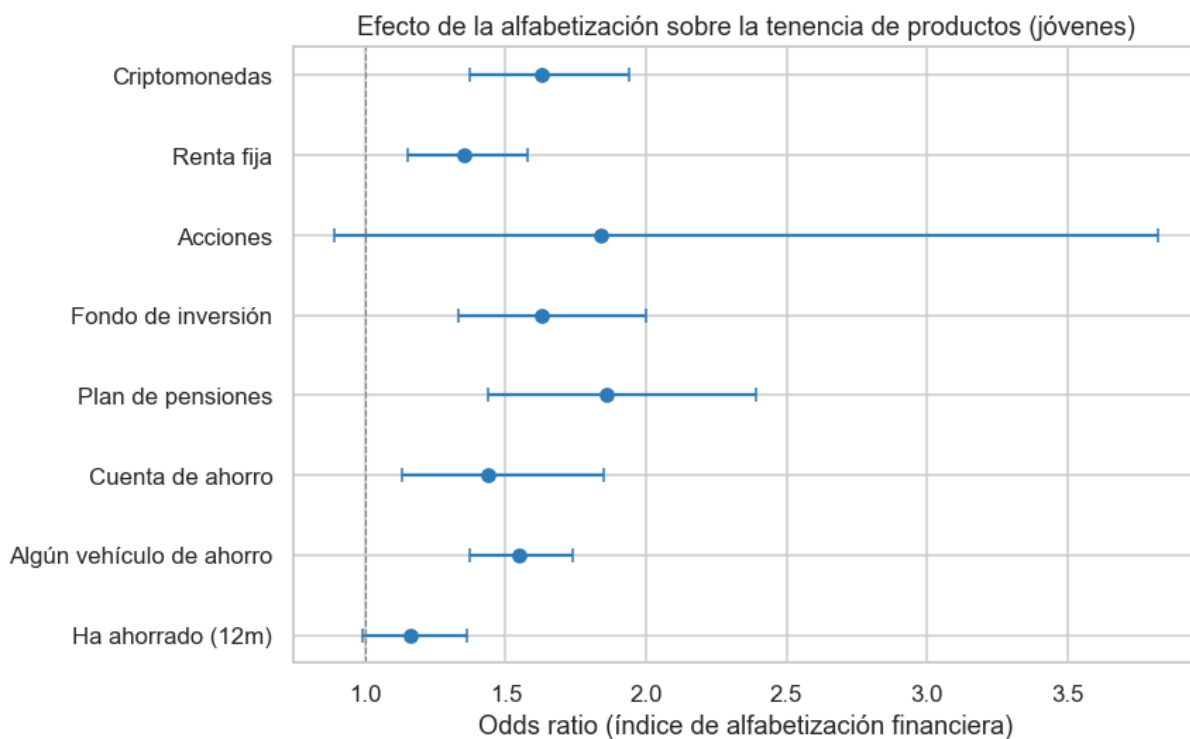
Out[22]:

```

	Producto	OR	IC 95% bajo	IC 95% alto	p-valor	N
0	Ha ahorrado (12m)	1.16	0.99	1.36	0.0757	1899
1	Algún vehículo de ahorro	1.55	1.37	1.74	0.0000	1899
2	Cuenta de ahorro	1.44	1.13	1.85	0.0034	1899
3	Plan de pensiones	1.86	1.44	2.39	0.0000	1899
4	Fondo de inversión	1.63	1.33	2.00	0.0000	1899

5	Acciones	1.84	0.89	3.82	0.0992	1899
6	Renta fija	1.35	1.15	1.58	0.0002	1899
7	Criptomonedas	1.63	1.37	1.94	0.0000	1899

```
In [23]: fig, ax = plt.subplots(figsize=(8, 5))
y_pos = np.arange(len(tabla_or))
ax.errorbar(
    tabla_or["OR"], y_pos,
    xerr=[tabla_or["OR"] - tabla_or["IC 95% bajo"],
          tabla_or["IC 95% alto"] - tabla_or["OR"]],
    fmt="o", color="#2b7bb9", capsize=3
)
ax.axvline(1, color="grey", lw=0.8, ls="--")
ax.set_yticks(y_pos)
ax.set_yticklabels(tabla_or["Producto"])
ax.set_xlabel("Odds ratio (índice de alfabetización financiera)")
ax.set_title("Efecto de la alfabetización sobre la tenencia de productos (jóvenes)")
plt.tight_layout()
plt.show()
```



Segmentación de perfiles

Aplico un análisis de conglomerados con k-medias para agrupar a los jóvenes según su alfabetización, su edad, el número de productos que poseen y su hábito de ahorro. Estandarizo previamente las variables.

```
In [24]: # Variables para el clustering
vars_cluster = ["indice_alfab", "age", "n_prod_ahorro", "ahorra"]
```

```
vars_cluster = [v for v in vars_cluster if v in df_joven.columns]

X = df_joven[vars_cluster].dropna()
print(f"Observaciones para clustering: {len(X):,}")

scaler = StandardScaler()
X_std = scaler.fit_transform(X)
```

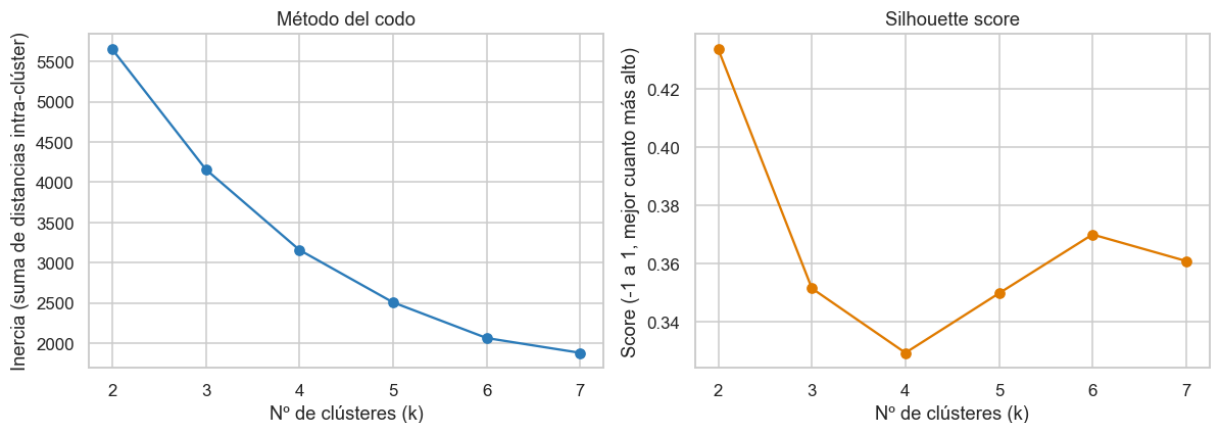
Observaciones para clustering: 1,899

Determino el número de grupos a partir del método del codo y del coeficiente de silueta.

```
In [25]: # Selección del nº óptimo de clústeres
inercias, silhouettes = [], []
ks = range(2, 8)
for k in ks:
    km = KMeans(n_clusters=k, random_state=RANDOM_STATE, n_init=10).fit(X_std)
    inercias.append(km.inertia_)
    silhouettes.append(silhouette_score(X_std, km.labels_))

fig, ax = plt.subplots(1, 2, figsize=(11, 4))
ax[0].plot(list(ks), inercias, marker="o", color="#2b7bb9")
ax[0].set_title("Método del codo")
ax[0].set_xlabel("Nº de clústeres (k)")
ax[0].set_ylabel("Inercia (suma de distancias intra-clúster)")

ax[1].plot(list(ks), silhouettes, marker="o", color="#e07b00")
ax[1].set_title("Silhouette score")
ax[1].set_xlabel("Nº de clústeres (k)")
ax[1].set_ylabel("Score (-1 a 1, mejor cuanto más alto)")
plt.tight_layout()
plt.show()
```



Fijo finalmente tres conglomerados, que ofrecen una segmentación interpretable, y examino el perfil medio de cada uno.

```
In [26]: K = 3

km = KMeans(n_clusters=K, random_state=RANDOM_STATE, n_init=10).fit(X_std)
X_clust = X.copy()
X_clust["cluster"] = km.labels_
```

```
# Perfil medio de cada clúster (sobre las variables originales, no estandarizadas)
perfil = X_clust.groupby("cluster").mean().round(2)
perfil["n"] = X_clust.groupby("cluster").size()
perfil["%"] = (perfil["n"] / perfil["n"].sum() * 100).round(1)
perfil
```

Out[26]:

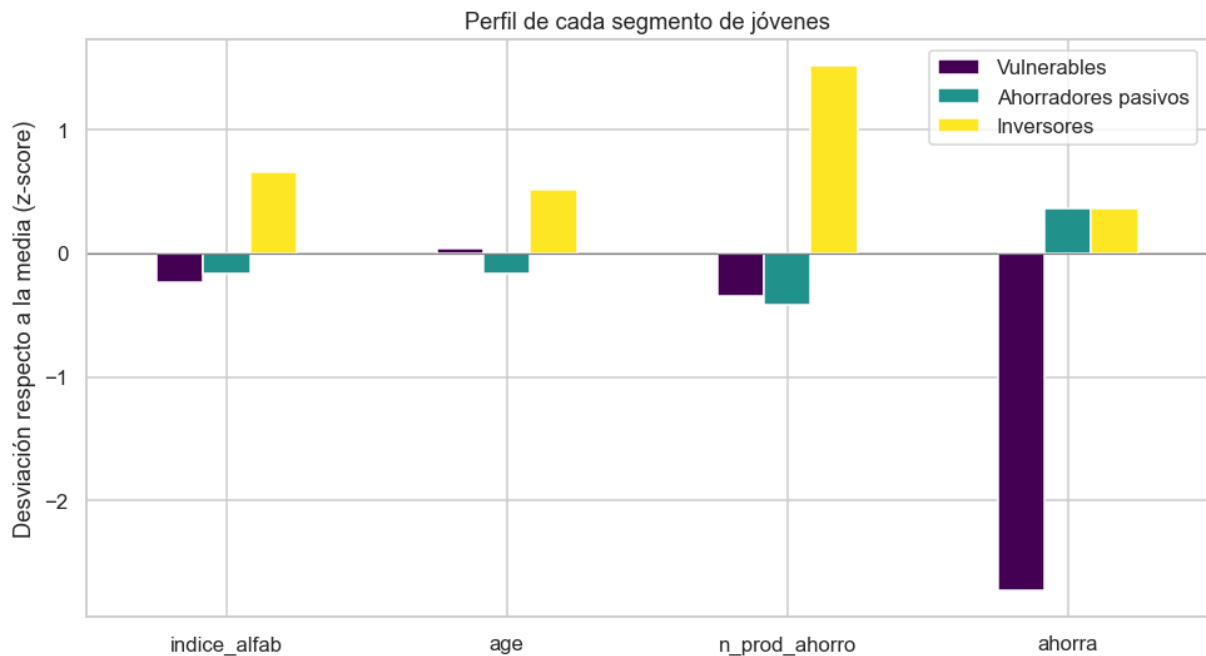
	indice_alfab	age	n_prod_ahorro	ahorra	n	%
cluster						
0	1.35	26.25	0.16	0.0	225	11.8
1	1.41	25.24	0.10	1.0	1270	66.9
2	2.19	28.57	1.63	1.0	404	21.3

In [27]:

```
# Perfil de cada clúster en valores estandarizados (z-score)
# Más claro para el TFG que la proyección PCA.
perfil_medias = X_clust.groupby("cluster")[vars_cluster].mean()
perfil_z = (perfil_medias - X[vars_cluster].mean()) / X[vars_cluster].std()

# Ajustar el mapeo según el orden de la tabla 'perfil' de arriba
nombres = {0: "Vulnerables", 1: "Ahorradores pasivos", 2: "Inversores"}
perfil_z.index = [nombres.get(i, f"Clúster {i}") for i in perfil_z.index]

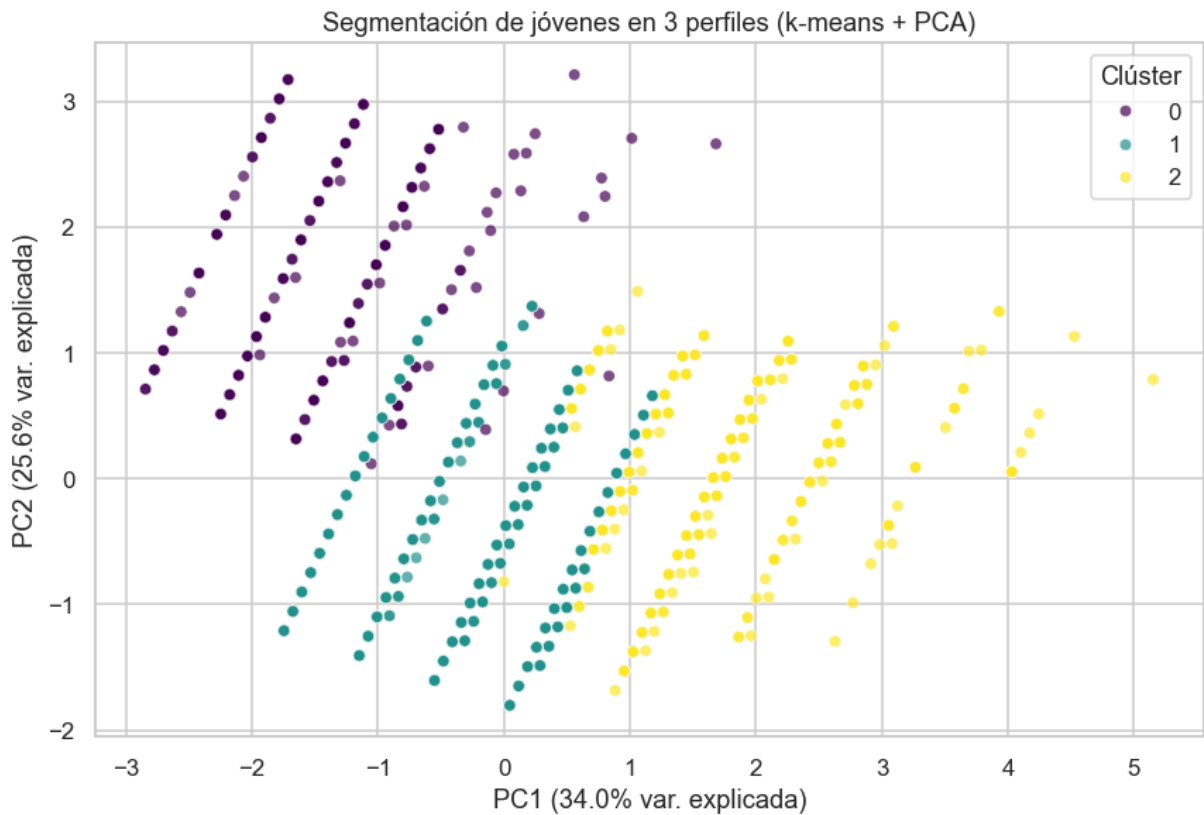
fig, ax = plt.subplots(figsize=(9, 5))
perfil_z.T.plot(kind="bar", ax=ax, colormap="viridis")
ax.axhline(0, color="grey", lw=0.8)
ax.set_title("Perfil de cada segmento de jóvenes")
ax.set_ylabel("Desviación respecto a la media (z-score)")
ax.set_xlabel("")
ax.legend(title="")
plt.xticks(rotation=0)
plt.tight_layout()
plt.show()
```



Por último proyecto los grupos sobre las dos primeras componentes principales para comprobar visualmente que ocupan regiones diferenciadas.

```
In [28]: pca = PCA(n_components=2, random_state=RANDOM_STATE)
proy = pca.fit_transform(X_std)

fig, ax = plt.subplots(figsize=(8, 5.5))
sns.scatterplot(
    x=proy[:, 0], y=proy[:, 1],
    hue=km.labels_, palette="viridis",
    s=30, alpha=0.7, ax=ax,
)
ax.set_xlabel(f"PC1 ({pca.explained_variance_ratio_[0]*100:.1f}% var. explicada)")
ax.set_ylabel(f"PC2 ({pca.explained_variance_ratio_[1]*100:.1f}% var. explicada)")
ax.set_title(f"Segmentación de jóvenes en {K} perfiles (k-means + PCA)")
ax.legend(title="Clúster")
plt.tight_layout()
plt.show()
```



Análisis de robustez

Para comprobar la solidez de los resultados repito el análisis bajo tres variaciones: aplicando los pesos muestrales, tratando el "no sabe" como categoría separada y restringiendo la cohorte a los 18 a 29 años.

En primer lugar incorporo el peso muestral de la encuesta, que reconstruye la población residente de 18 a 79 años, y reestimo los descriptivos y los modelos principales.

```
In [29]: # Identificación y comprobación de la variable de peso muestral

# 1) Cargar etiquetas del fichero .dta
with pd.io.stata.StataReader(RUTA_ECF_2021) as reader:
    var_labels = reader.variable_labels()

# 2) Buscar candidatas a variable de peso
candidatas = [c for c in df.columns if any(
    k in c.lower() for k in ["peso", "factor", "weight", "ponder", "w_", "_w"]
)]
print("Candidatas a peso:", candidatas)
for v in candidatas:
    print(f" {v}: {var_labels.get(v, '(sin etiqueta)')}")

# 3) Inspeccionar la variable 'weight'
print("\nResumen de 'weight':")
print(df["weight"].describe())
```

```
print(f"\nValores faltantes: {df['weight'].isna().sum()}")
print(f"Suma total de pesos: {df['weight'].sum():.0f}")
```

Candidatas a peso: ['weight']

weight: weight: peso

Resumen de 'weight':

```
count    7764.000000
mean     4639.768555
std       2990.278809
min       420.189911
25%      2446.758545
50%      3852.094482
75%      6402.205566
max      19411.416016
Name: weight, dtype: float64
```

Valores faltantes: 0

Suma total de pesos: 36,023,164

In [30]: *# Análisis de robustez 1: descriptivos ponderados*

```
def media_ponderada(serie, pesos):
    """Media ponderada que ignora NaN."""
    valida = serie.notna() & pesos.notna()
    return np.average(serie[valida], weights=pesos[valida])

# Distribución del índice 0-3 en jóvenes, ponderada
peso_joven = df_joven["weight"]
print("Distribución del índice de alfabetización en jóvenes (ponderada, %):")
for k in [0, 1, 2, 3]:
    pct = (df_joven.loc[df_joven["indice_alfab"] == k, "weight"].sum()
           / peso_joven.sum() * 100)
    print(f" {k} respuestas correctas: {pct:.1f}%")

# Porcentaje con alfabetización plena (las tres correctas)
pct_plena = media_ponderada(df_joven["alfab_plena"], peso_joven) * 100
print(f"\n% con alfabetización plena (3 correctas), ponderado: {pct_plena:.1f}%")
print("(Cifra publicada por el BdE para jóvenes 18-34: en torno al 19%)")

# Comparación con la cifra no ponderada para ver el efecto
pct_plena_sin = df_joven["alfab_plena"].mean() * 100
print(f"\nPara contraste, sin ponderar: {pct_plena_sin:.1f}%")
```

Distribución del índice de alfabetización en jóvenes (ponderada, %):

```
0 respuestas correctas: 14.0%
1 respuestas correctas: 34.6%
2 respuestas correctas: 34.3%
3 respuestas correctas: 17.2%
```

% con alfabetización plena (3 correctas), ponderado: 17.2%

(Cifra publicada por el BdE para jóvenes 18-34: en torno al 19%)

Para contraste, sin ponderar: 18.2%

In [31]: *# Análisis de robustez 2: regresión lineal con pesos*

```

formula = f"indice_alfab ~ {controles}"

modelo_lineal_pond = smf.wls(
    formula,
    data=df_joven,
    weights=df_joven["weight"]
).fit()

print(modelo_lineal_pond.summary())

```

WLS Regression Results

```

=====
Dep. Variable:      indice_alfab    R-squared:      0.096
Model:              WLS             Adj. R-squared:  0.091
Method:             Least Squares   F-statistic:    18.23
Date:               Wed, 17 Jun 2026 Prob (F-statistic): 5.47e-35
Time:               19:47:21        Log-Likelihood: -2667.9
No. Observations:   1899           AIC:            5360.
Df Residuals:       1887           BIC:            5426.
Df Model:           11
Covariance Type:    nonrobust
=====

```

```

=====
                                coef    std err          t      P>|t|      [0.025
0.975]
-----
-----
Intercept                    1.0210      0.251      4.065      0.000      0.528
1.514
C(sexo)[T.Mujer]             -0.3152      0.042     -7.567      0.000     -0.397
-0.234
C(educ_level)[T.2]           0.3996      0.056      7.195      0.000      0.291
0.509
C(educ_level)[T.3]           0.6088      0.056     10.817      0.000      0.498
0.719
C(labor_status)[T.2.0]       -0.2350      0.084     -2.810      0.005     -0.399
-0.071
C(labor_status)[T.3.0]       -0.2435      0.097     -2.499      0.013     -0.435
-0.052
C(labor_status)[T.4.0]       -1.4886      0.581     -2.563      0.010     -2.628
-0.349
C(labor_status)[T.5.0]       -0.0116      0.094     -0.123      0.902     -0.196
0.173
C(subgrupo_edad)[T.25-34]    0.1096      0.053      2.076      0.038      0.006
0.213
C(riqueza_in)[T.1]           0.3923      0.238      1.645      0.100     -0.075
0.860
C(riqueza_in)[T.2]           0.3580      0.238      1.506      0.132     -0.108
0.824
C(riqueza_in)[T.3]           0.4131      0.238      1.735      0.083     -0.054
0.880
=====

```

```

=====
Omnibus:      0.929    Durbin-Watson:      1.971
Prob(Omnibus): 0.629    Jarque-Bera (JB):      0.867
Skew:         -0.049    Prob(JB):              0.648
Kurtosis:     3.036    Cond. No.              46.9
=====

```

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [32]: *# Análisis de robustez 3: regresión logística con pesos (versión corregida con GLM)*

```
def logit_producto_ponderado(df, dep_var, controles):
    """Logit ponderado vía GLM con familia binomial.
    smf.logit no admite freq_weights directamente; smf.glm sí.
    El resultado es estadísticamente equivalente al logit estándar."""
    formula = f"{dep_var} ~ indice_alfab + {controles}"
    return smf.glm(
        formula,
        data=df,
        family=sm.families.Binomial(),
        freq_weights=df["weight"].values
    ).fit()

resultados_pond = {}
for v in vars_dep:
    if v in df_joven.columns:
        try:
            resultados_pond[v] = logit_producto_ponderado(df_joven, v, controles)
        except Exception as e:
            print(f"Error en {v}: {e}")

# Tabla comparativa: OR sin pesos vs con pesos
tabla_comp = []
for v in resultados:
    if v in resultados_pond and "indice_alfab" in resultados[v].params.index:
        or_sin = np.exp(resultados[v].params["indice_alfab"])
        or_con = np.exp(resultados_pond[v].params["indice_alfab"])
        p_sin = resultados[v].pvalues["indice_alfab"]
        p_con = resultados_pond[v].pvalues["indice_alfab"]
        tabla_comp.append({
            "Producto": v,
            "OR sin pesos": round(or_sin, 2),
            "p sin pesos": round(p_sin, 4),
            "OR con pesos": round(or_con, 2),
            "p con pesos": round(p_con, 4),
        })

pd.DataFrame(tabla_comp)
```

Out[32]:

	Producto	OR sin pesos	p sin pesos	OR con pesos	p con pesos
0	ahorra	1.16	0.0757	1.15	0.0
1	t_alahoc	1.55	0.0000	1.53	0.0
2	t_cuenta_ahorro	1.44	0.0034	1.53	0.0
3	t_plan_pensiones	1.86	0.0000	2.02	0.0
4	t_fondo_inv	1.63	0.0000	1.51	0.0

5	t_acciones	1.84	0.0992	1.67	0.0
6	t_renta_fija	1.35	0.0002	1.36	0.0
7	t_cripto	1.63	0.0000	1.61	0.0

Después construyo un índice alternativo que trata el "no sabe" como respuesta no válida en lugar de como error, y comparo la brecha de género bajo ambas codificaciones.

```
In [33]: # Análisis de robustez 4: tratamiento del "no sé" como categoría separada
# (Conexión con Hospido, Iriberry y Machelett, 2024)

# Construir versiones alternativas de las tres preguntas:
# 1 si correcta, 0 si incorrecta, NaN si "no sé" o "no contesta"
def correcta_o_nan(serie, valores_correctos):
    """1 si la respuesta es correcta; 0 si es incorrecta; NaN si NS/NC."""
    resultado = serie.isin(valores_correctos).astype(float)
    resultado[serie.isin(CODIGOS_NSNC)] = np.nan
    return resultado

df["correct_e0600_alt"] = correcta_o_nan(df["e0600"], [3, 4])
df["correct_e0900_alt"] = correcta_o_nan(df["e0900"], [1])
df["correct_e1003_alt"] = correcta_o_nan(df["e1003"], [1])

# Índice alternativo: proporción de aciertos sobre preguntas respondidas
preguntas_alt = ["correct_e0600_alt", "correct_e0900_alt", "correct_e1003_alt"]
df["indice_alfab_alt"] = df[preguntas_alt].mean(axis=1) # ignora NaN automáticamente

# Replicar también en la submuestra joven
df_joven["indice_alfab_alt"] = df.loc[df_joven.index, "indice_alfab_alt"]

print("Comparación de los dos índices en jóvenes:")
print(f"Índice original (0-3, NS = incorrecto): media = {df_joven['indice_alfa']
print(f"Índice alternativo (proporción, NS = NaN): media = {df_joven['indice_alfab']
# El recuento de jóvenes con algún "no sé" se calcula correctamente en el bloque 10
```

Comparación de los dos índices en jóvenes:

Índice original (0-3, NS = incorrecto): media = 1.572

Índice alternativo (proporción, NS = NaN): media = 0.578

```
In [34]: # Comparar la brecha de género con cada uno de los dos índices

# Modelo con el índice original (el que ya tenías)
m_original = smf.ols(f"indice_alfab ~ {controles}", data=df_joven).fit()

# Modelo con el índice alternativo (escalado a 0-3 para que sea comparable)
df_joven["indice_alfab_alt_x3"] = df_joven["indice_alfab_alt"] * 3
m_alt = smf.ols(f"indice_alfab_alt_x3 ~ {controles}", data=df_joven).fit()

# Coeficiente de "Mujer" en ambos modelos
print("Brecha de género (coeficiente de C(sexo)[T.Mujer]) en escala 0-3:")
print(f"Índice original (NS = incorrecto): {m_original.params['C(sexo)[T.Mujer]']
      f"(p = {m_original.pvalues['C(sexo)[T.Mujer]']:.4f})")
print(f"Índice alternativo (NS = NaN): {m_alt.params['C(sexo)[T.Mujer]']:.3f})
```

```

f"(p = {m_alt.pvalues['C(sexo)[T.Mujer]']:.4f})")

# Cambio porcentual en la brecha
brecha_orig = m_original.params["C(sexo)[T.Mujer]"]
brecha_alt = m_alt.params["C(sexo)[T.Mujer]"]
cambio_pct = (brecha_alt - brecha_orig) / abs(brecha_orig) * 100
print(f"\nCambio en la brecha: {cambio_pct:+.1f}% "
      f"({'se reduce' if abs(brecha_alt) < abs(brecha_orig) else 'se mantiene o amp

```

Brecha de género (coeficiente de C(sexo)[T.Mujer]) en escala 0-3:

Índice original (NS = incorrecto): -0.348 (p = 0.0000)

Índice alternativo (NS = NaN): -0.241 (p = 0.0000)

Cambio en la brecha: +30.6% (se reduce)

```

In [35]: con_ns = df.loc[df_joven.index, preguntas_alt].isna().any(axis=1).sum()
print(f"Jóvenes con al menos un 'no sé' en las Big Three: {con_ns} ({con_ns/len(df_

```

Jóvenes con al menos un 'no sé' en las Big Three: 446 (23.5%)

Por último restrinjo la cohorte a los 18 a 29 años y verifico que los coeficientes y los odds ratios principales se mantienen.

```

In [36]: # Análisis de robustez 5: estabilidad ante el corte de edad

# Submuestra restringida a 18-29
df_1829 = df_joven[df_joven["age"].between(18, 29)].copy()
print(f"Submuestra 18-29: {len(df_1829):,} observaciones")

# Reestimar el modelo lineal y los logit centrales sobre la submuestra
m_1829 = smf.ols(f"indice_alfab ~ {controles}", data=df_1829).fit()

print("\nCoeficiente de 'Mujer' en distintas submuestras:")
print(f"  18-34 (referencia): {m_original.params['C(sexo)[T.Mujer]']:.3f} "
      f"(p = {m_original.pvalues['C(sexo)[T.Mujer]']:.4f})")
print(f"  18-29 (restringido): {m_1829.params['C(sexo)[T.Mujer]']:.3f} "
      f"(p = {m_1829.pvalues['C(sexo)[T.Mujer]']:.4f})")

print("\nCoeficiente de 'Universitario' en distintas submuestras:")
print(f"  18-34 (referencia): {m_original.params['C(edu_level)[T.3]']:.3f} "
      f"(p = {m_original.pvalues['C(edu_level)[T.3]']:.4f})")
print(f"  18-29 (restringido): {m_1829.params['C(edu_level)[T.3]']:.3f} "
      f"(p = {m_1829.pvalues['C(edu_level)[T.3]']:.4f})")

# Logit central (algún vehículo de ahorro) en ambas submuestras
m_log_1829 = smf.glm(
    f"t_alahoc ~ indice_alfab + {controles}",
    data=df_1829,
    family=sm.families.Binomial()
).fit()

or_1834 = np.exp(resultados["t_alahoc"].params["indice_alfab"])
or_1829 = np.exp(m_log_1829.params["indice_alfab"])
print(f"\nOR del índice de alfabetización para 'algún vehículo de ahorro':")
print(f"  18-34 (referencia): OR = {or_1834:.2f}")
print(f"  18-29 (restringido): OR = {or_1829:.2f}")

```

Submuestra 18-29: 1,322 observaciones

Coeficiente de 'Mujer' en distintas submuestras:

18-34 (referencia): -0.348 (p = 0.0000)

18-29 (restringido): -0.367 (p = 0.0000)

Coeficiente de 'Universitario' en distintas submuestras:

18-34 (referencia): 0.628 (p = 0.0000)

18-29 (restringido): 0.597 (p = 0.0000)

OR del índice de alfabetización para 'algún vehículo de ahorro':

18-34 (referencia): OR = 1.55

18-29 (restringido): OR = 1.55