

# Metodología de nuestro modelo de estimación electoral

---

Tarek Jaziri Arjona

---

Abril 2026



# Índice

<b>1</b>	<b>Introducción</b>	<b>1</b>
<b>2</b>	<b>Resumen Ejecutivo</b>	<b>4</b>
<b>3</b>	<b>La ponderación de la muestra</b>	<b>7</b>
3.1	Una estrategia en dos etapas . . . . .	10
<b>4</b>	<b>El ajuste del recuerdo de voto</b>	<b>11</b>
4.1	Cómo imputamos el recuerdo de voto . . . . .	13
4.2	Validación y cierre del ciclo de ponderación . . . . .	15
<b>5</b>	<b>Imputando la intención de voto a los indecisos</b>	<b>18</b>
5.1	Qué cambia respecto a la imputación del recuerdo . . . . .	19
5.2	El destino de los indecisos tras la imputación . . . . .	21
<b>6</b>	<b>¿Quién irá a votar?</b>	<b>22</b>
6.1	Qué información utiliza el modelo . . . . .	24
6.2	De la probabilidad al voto: cómo decidimos quién votará . . . . .	25
<b>7</b>	<b>La estimación final</b>	<b>29</b>
7.1	Cómo calculamos la estimación de voto . . . . .	29
7.2	Cuantificación de la incertidumbre . . . . .	31
<b>8</b>	<b>Conclusión</b>	<b>32</b>
<b>A</b>	<b>Validación de los modelos de clasificación</b>	<b>35</b>
A.1	Modelo de recuerdo de voto . . . . .	35
A.2	Modelo de intención de voto . . . . .	36
A.3	Capacidad de discriminación: curvas ROC . . . . .	37
A.4	Regularización y sobreajuste: curvas de aprendizaje . . . . .	39

# 1. Introducción

---

Imagina que vas a tirar con arco, pero las condiciones no juegan a tu favor. Hay niebla y la diana queda lejos, difusa. Además, sopla un viento fuerte desde el oeste: sabes que desviará la flecha, aunque los meteorólogos avisaron el día antes de que podría aflojar... o incluso girar y venir del este súbitamente. Aun así, tienes que decidir: ¿Apuntas un poco a la izquierda para compensar? ¿Cuánto? ¿Y si el viento cambia justo cuando sueltas la cuerda del arco?

Hacer estimaciones electorales se parece un poco a eso: no tiras a ciegas, pero tampoco tienes una foto perfecta. Trabajas con información parcial, con ruido y con fuerzas que empujan en direcciones distintas. Y el reto no es solo “dar en el centro”, sino explicar con honestidad cuánto puede desviarse la flecha y por qué.

Estimar el resultado de unas elecciones nunca es sencillo, pero hacerlo en España presenta dificultades propias. El sistema de partidos se ha fragmentado enormemente en la última década: hemos pasado de un casi-bipartidismo relativamente estable a un escenario con cuatro, cinco o más partidos compitiendo por escaños en el Congreso. Esa fragmentación no es uniforme: el mapa territorial añade capas de complejidad, con partidos regionalistas y nacionalistas que dominan en sus comunidades pero apenas existen fuera de ellas.

A esto se suma una volatilidad electoral que no deja de crecer. En pocos años hemos visto la irrupción y caída de Ciudadanos, el ascenso de Vox, la reconfiguración de la izquierda entre Podemos y Sumar, y la aparición de nuevas fuerzas como Se Acabó La Fiesta. Cada ciclo electoral parece reescribir las reglas del anterior, y para un modelo de estimación, esto significa que los patrones del pasado son una guía útil pero insuficiente: el electorado español se mueve, y se mueve rápido.

Pero las dificultades no terminan en el panorama político: empiezan también en los propios datos. Toda encuesta es una aproximación imperfecta a la realidad. La muestra nunca es un espejo exacto de la población: ciertos perfiles responden más que otros, lo que introduce sesgos sistemáticos que hay que corregir. Hay encuestados que no recuerdan (o no quieren decir) a quién votaron en las últimas elecciones (Urquiza, 2005), una información que resulta clave para calibrar el modelo. Otros no saben aún qué harán en las próximas y asignarles un comportamiento probable es una de las decisiones más influyentes de toda la estimación. A todo esto se añade la cuestión de la participación: no basta con saber a quién votaría cada persona si fuera a votar; hay que estimar quién irá realmente a las ur-

nas, un ejercicio que se complica aún más porque el propio modo de administración de la encuesta (telefónica, presencial, online) condiciona a quién se llega y a quién no, introduciendo sesgos específicos que hay que tener en cuenta (Groves et al., 2009). Cada una de estas capas de incertidumbre se acumula, y gestionarlas con rigor es precisamente lo que distingue una estimación cuidadosa de un simple recuento de respuestas.

En este contexto, las encuestas electorales juegan un papel central en el debate público. Marcan expectativas, condicionan estrategias de partidos y, en no pocas ocasiones, generan relatos que luego la realidad no confirma: el supuesto *sorpasso* de Podemos al PSOE en 2015, las mayorías absolutas de la derecha que varias encuestas anticipaban en 2023... Son solo un par de ejemplos de cómo unas estimaciones mal calibradas o mal comunicadas pueden distorsionar el debate democrático (Shirani-Mehr et al., 2018; Kennedy et al., 2018).

Y sin embargo, el proceso detrás de esas cifras sigue siendo, en gran medida, opaco. En España, todavía es raro ver publicados los microdatos de las encuestas electorales. Y es más raro todavía acceder a los detalles de la *cocina* exacta que emplea cada encuestadora (Ferrándiz y Camas García, 2019; Maldita.es, 2019): los ajustes, ponderaciones y decisiones metodológicas que transforman las respuestas crudas de una encuesta en la estimación de voto que termina en un titular.

Desde More in Common creemos que, si las encuestas van a influir inevitablemente en las expectativas de los ciudadanos, en la cobertura de los medios y en las estrategias de los partidos, entonces la sociedad tiene derecho a saber cómo se construyen esas cifras. Por eso hemos decidido empezar a publicar nuestras propias estimaciones electorales de forma periódica, un esfuerzo que, además, tiene otros objetivos:

1. En primer lugar, queremos que actores que tradicionalmente no tienen el tiempo, los recursos o los conocimientos para ello tengan acceso a la misma información que normalmente solo está en manos de unas pocas agencias demoscópicas o de partidos políticos. Creemos que la opinión pública no debería ser un recurso reservado a partidos políticos y grandes medios, sino un bien común al alcance de cualquiera que quiera entender cómo piensa y se mueve la sociedad.
2. En segundo lugar, contar con un modelo de estimación propio nos permitirá profundizar en nuestro conocimiento y análisis de las dinámicas sociales en España. La volatilidad electoral mencionada más arriba tiene sus raíces en procesos de polarización y fractura social, y conocer y estudiar esos procesos resulta fundamental para

cumplir nuestra misión como organización.

3. Por último, queremos contribuir a la transparencia y al debate en el sector. Contamos con la ventaja de que nuestro modelo electoral no está realizado para ningún medio, organización ni partido en concreto, por lo que tenemos total libertad para tomar las decisiones que queramos y explicarlas de manera pública y con un nivel de detalle y transparencia inédito en nuestro país. Los microdatos de nuestras estimaciones serán siempre públicos: estamos totalmente seguros de la calidad de nuestros datos y, al mismo tiempo, totalmente abiertos a recibir críticas y aportaciones sobre cómo mejorar nuestro modelo electoral, que desde el principio hemos concebido como un esfuerzo dinámico.

En este documento explicamos el método completo que empleamos para llegar hasta nuestras estimaciones electorales: cada decisión, cada supuesto, cada limitación.

Es importante ser claros desde el principio sobre lo que este modelo hace y lo que no hace. Una estimación electoral no es una predicción. No pretende adivinar qué ocurrirá el día de las elecciones, porque entre el momento de la encuesta y la votación pueden cambiar muchas cosas: campañas, eventos inesperados, decisiones de última hora... Lo que ofrece es una fotografía del presente: si las elecciones se celebrasen hoy, con la información que tenemos hoy, estos serían los resultados más probables y este sería el margen de incertidumbre. Su valor no está en acertar el resultado al decimal, sino en ayudar a entender las dinámicas del electorado: quién se mueve, hacia dónde, por qué, y con qué grado de certeza podemos afirmarlo. Una encuesta que admite lo que no sabe es más valiosa que una que finge precisión.

El principio que guía nuestro enfoque es dejar que los datos hablen en la mayor medida posible. En toda estimación electoral hay decisiones que alguien tiene que tomar: qué variables utilizar, cómo tratar a los indecisos, dónde poner el umbral de participación, etc. Nuestra aproximación es la siguiente: en cada punto donde el modelo necesita una decisión, hemos buscado un criterio que emerja de los propios datos en lugar de imponerlo desde fuera; no siempre es posible (y señalaremos los casos en que no lo es), pero la orientación general del modelo es minimizar las decisiones arbitrarias y maximizar las que están respaldadas empíricamente.

En definitiva, en las páginas que siguen, describimos paso a paso el modelo de estimación electoral que hemos desarrollado desde More in Common España. Explicamos por qué tomamos cada decisión, qué alternativas consideramos y qué limitaciones tiene el en-

foque elegido. Lo hacemos en un lenguaje que aspira a ser accesible sin sacrificar el rigor: los conceptos técnicos se introducen cuando son necesarios, y se acompañan siempre de una explicación intuitiva.

Por último, y antes de pasar a la explicación del modelo, nos gustaría también aclarar que para llegar a él hemos consultado numerosas fuentes bibliográficas y académicas referenciadas en este documento, hablado con distintos expertos en estimaciones electorales para recoger sus perspectivas, e inspirado nuestro trabajo en modelos y experiencias anteriores, entre las cuales queremos destacar dos: el trabajo realizado por los equipos de investigación de More in Common en otros países como el Reino Unido y el modelo desarrollado en su día por Gonzalo Rivero y Jordi Muñoz durante su etapa en el Centre d'Estudis d'Opinió de la Generalitat de Catalunya (2022–2024).

Agradecemos a Blanca Domenech Pelleja su colaboración en la revisión metodológica, la validación de resultados y la discusión crítica del modelo. Sus aportaciones como segundo par de ojos han contribuido a la robustez del trabajo presentado en este report.

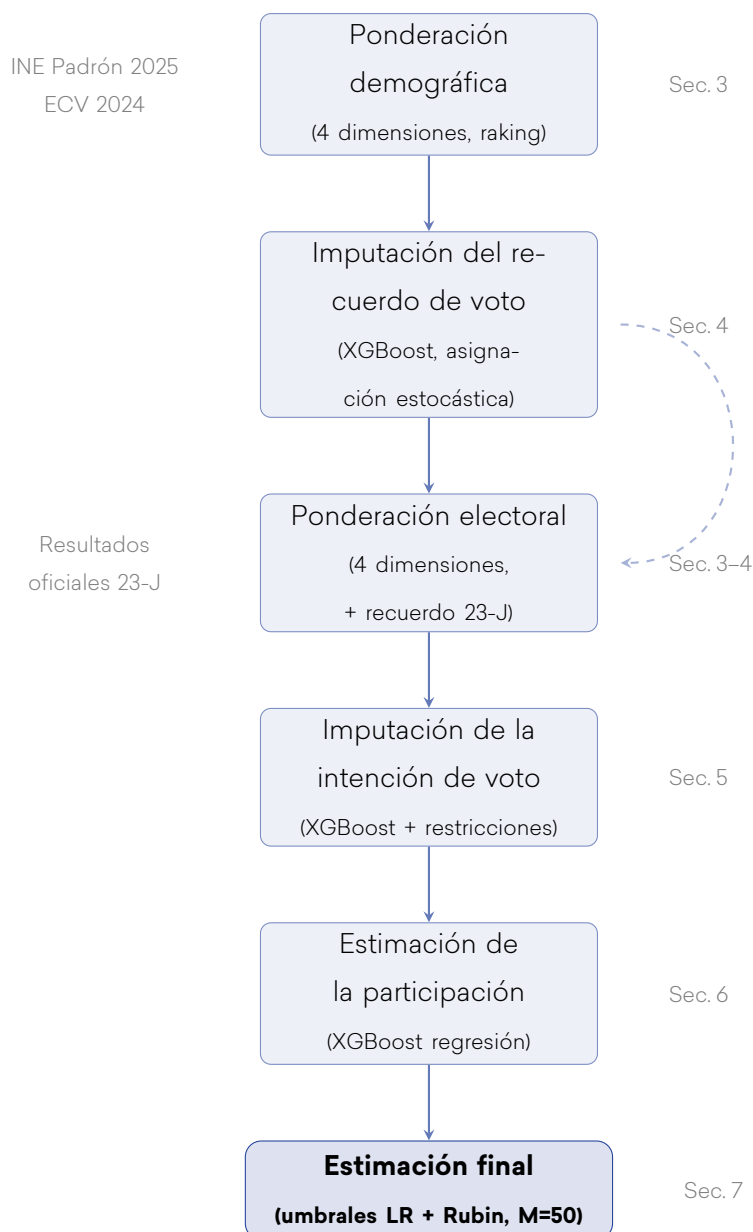
## 2. Resumen Ejecutivo

---

Este informe describe la metodología completa del modelo de estimación electoral de More in Common España. Lo publicamos junto con el código y los datos porque creemos firmemente en la transparencia: las cifras que marcan el debate público deben venir acompañadas siempre de su método, sus supuestos y sus limitaciones. Con todo ello, nuestro objetivo es muy sencillo: democratizar la investigación por encuesta de calidad, poniéndola al alcance de la sociedad civil, los investigadores y cualquier ciudadano que quiera entender cómo piensa y se mueve el electorado.

### A. El proceso en seis pasos

La Figura 1 resume el proceso completo que seguimos en nuestro modelo de estimación electoral. Cada etapa se describe en detalle en la sección correspondiente del informe.



**Figura 1.** Diagrama del proceso de estimación electoral. Cada caja representa una etapa del modelo, con la referencia a la sección del informe donde se describe. La flecha discontinua marca el ciclo de ponderación en dos etapas.

1. **Ponderación demográfica** (Sec. 3). Corregimos los desequilibrios de la muestra mediante raking en cuatro dimensiones (género por edad, Comunidad Autónoma, tamaño del municipio y renta), utilizando datos oficiales del INE como referencia.
2. **Imputación del recuerdo de voto** (Sec. 4). Los encuestados que no declaran a quién votaron en 2023 se imputan mediante un modelo XGBoost que aprende de quienes sí responden. La asignación es estocástica: en lugar de asignar el partido más probable, se sortea según la distribución de probabilidades predicha, preservando la variabilidad real del electorado.

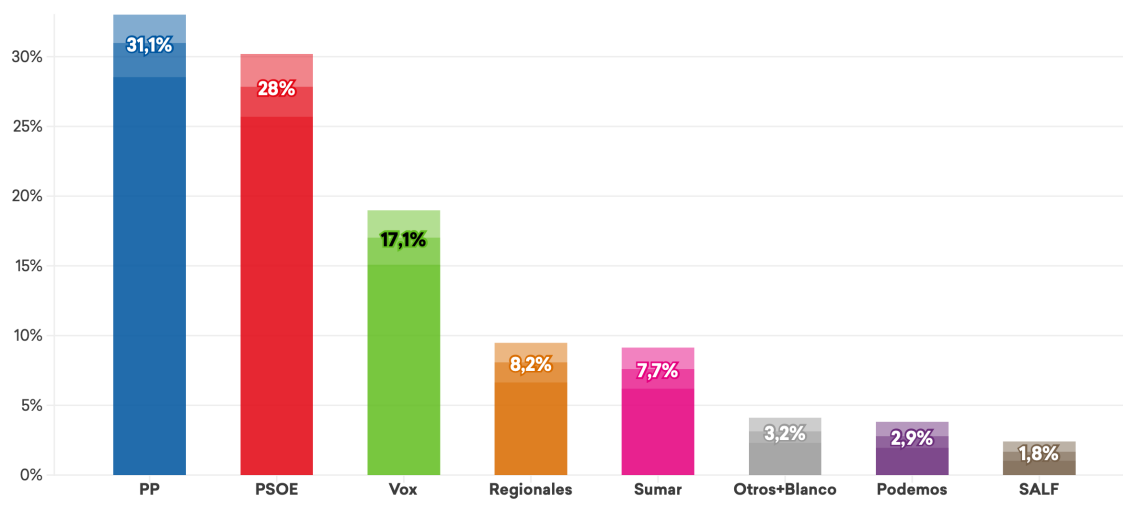
3. **Ponderación electoral** (Sec. 3–4). Con el recuerdo completo, añadimos una quinta dimensión de raking calibrada contra los resultados oficiales del 23-J, corrigiendo posibles sesgos partidistas de la muestra.
4. **Imputación de la intención de voto** (Sec. 5). Los indecisos se imputan con un segundo modelo XGBoost que incorpora el recuerdo ya imputado como predictor. Se aplican restricciones basadas en los vetos declarados por cada encuestado y una corrección de sesgo calibrada contra 2023.
5. **Estimación de la participación** (Sec. 6). Un modelo XGBoost de regresión estima la probabilidad individual de votar. La clasificación como votante o abstencionista se realiza mediante umbrales diferenciados por grupo, derivados empíricamente de los datos a través de un cociente de verosimilitud.
6. **Estimación final y cuantificación de la incertidumbre** (Sec. 7). Las cuotas de voto se calculan combinando los pesos de la encuesta con la probabilidad de participación. La incertidumbre se cuantifica mediante imputación múltiple (50 re-sorteos) y las reglas de Rubin, que combinan el error muestral con el error de imputación en un único intervalo de confianza.

## B. Resultado principal

La Figura 2 muestra la estimación de voto resultante. Para cada partido se representan tres puntos: el central indica la estimación puntual y los extremos delimitan el intervalo de confianza al 95%. La estimación no es una predicción de unas elecciones futuras, sino una fotografía del estado actual de las preferencias del electorado, con sus incertidumbres explícitamente cuantificadas.

## Estimación de voto

Participación: **68,3%** [65,8-72,12%]



% sobre voto válido.

Intervalos de confianza al 95%.

La incertidumbre se ha calculado según las reglas de Rubin para imputación múltiple (M = 50). Trabajo de campo del 21 al 25 de abril de 2026. N=2547

**Figura 2.** Estimación de voto. Para cada partido se representan tres puntos: el central indica la estimación puntual y los extremos los límites del intervalo de confianza al 95 %, derivado de las reglas de Rubin (M=50). Las cifras se expresan como porcentaje del voto válido.

## C. Principios del modelo

Tres ideas vertebran el diseño del modelo electoral:

- **Transparencia.** Publicamos el informe metodológico completo y los microdatos de la encuesta. Cualquiera puede evaluar nuestras estimaciones y entender cómo se construyen.
- **Decisiones basadas en datos.** En cada punto donde el modelo necesita una decisión, hemos buscado un criterio empírico en lugar de imponer uno arbitrario.
- **Honestidad sobre la incertidumbre.** Todas las estimaciones se acompañan de intervalos de confianza que reflejan tanto el error muestral como el error de imputación.

## 3. La ponderación de la muestra

Ninguna encuesta es un espejo exacto de la población. Aunque el diseño muestral busca representar a la sociedad española, en la práctica siempre hay perfiles que responden más que otros. Los jóvenes suelen estar infrarrepresentados, ciertos tramos de renta participan

menos y algunas Comunidades Autónomas acaban con más o menos peso del que les corresponde. Si no se corrigen estos desequilibrios, las opiniones de los grupos sobre-representados pesan demasiado y las del resto, demasiado poco. El resultado sería una estimación sesgada antes incluso de haber empezado a modelar nada.

La ponderación es la herramienta que corrige esto. La idea es sencilla: asignar a cada encuestado un peso numérico que refleje a cuántas personas de la población real representa. Si en la muestra hay demasiados hombres de 55 a 64 años, sus respuestas recibirán un peso menor; si hay pocas mujeres de 18 a 24, las suyas recibirán un peso mayor. El objetivo es que, una vez aplicados los pesos, los márgenes de la muestra coincidan con los de la población real (Pew Research Center, [2018](#)).

El método que utilizamos para ello se llama *raking*, también conocido como ajuste iterativo proporcional. Es una técnica con una larga trayectoria en la estadística de encuestas, propuesta originalmente en el contexto del censo de Estados Unidos de 1940 (Deming y Stephan, [1940](#)) y formalizada posteriormente dentro del marco general de estimadores de calibración (Deville y Särndal, [1992](#)). Funciona de la siguiente manera: se definen varias dimensiones para las que conocemos la distribución real de la población. El algoritmo ajusta los pesos de forma iterativa, dimensión por dimensión, hasta que los márgenes de la muestra ponderada coinciden con los de la población en todas las dimensiones simultáneamente.

La ponderación de encuestas no es una tarea sencilla (Gelman, [2007](#)). Existen otros métodos de ponderación. La post-estratificación, por ejemplo, divide la muestra en celdas definidas por el cruce completo de todas las variables (género  $\times$  edad  $\times$  CCAA  $\times$  tamaño de municipio  $\times$  renta) y ajusta el peso de cada celda a su proporción real en la población. El problema es que, al cruzar muchas dimensiones, el número de celdas crece rápidamente y muchas quedan con muy pocos encuestados o directamente vacías, lo que produce pesos extremos e inestables (Lumley, [2010](#)). Otra alternativa son los métodos basados en propensión (*propensity score weighting*), que modelan la probabilidad de responder a la encuesta y ponderan en función inversa de esa probabilidad (Groves et al., [2009](#)). Estos métodos pueden ser útiles cuando se dispone de información individual sobre los no respondientes, algo que en nuestro caso no tenemos.

El raking ofrece un equilibrio práctico entre ambos extremos: permite incorporar múltiples dimensiones de ajuste sin necesidad de que la muestra tenga suficientes casos en cada combinación posible de categorías (por ejemplo, mujeres jóvenes de renta alta en

Extremadura). Basta con que los márgenes de cada dimensión por separado estén bien cubiertos. Es, además, el método más utilizado en la práctica de las encuestas electorales por su flexibilidad, su transparencia y su buen comportamiento con muestras de tamaño moderado (Pew Research Center, 2018).

Nuestro primer paso de ponderación utiliza cuatro dimensiones, cuyos valores de referencia provienen de fuentes oficiales del INE (Instituto Nacional de Estadística, 2025; Instituto Nacional de Estadística, 2024):

1. **Géneros por grupo de edad:** 12 celdas que cruzan dos géneros (hombre y mujer) con seis tramos de edad (18–24, 25–34, 35–44, 45–54, 55–64 y 65+).
2. **Comunidad Autónoma:** las 17 Comunidades Autónomas más las dos ciudades autónomas (Ceuta y Melilla).
3. **Tamaño del municipio:** seis categorías, desde municipios rurales hasta grandes ciudades.
4. **Ingresos netos del hogar:** cinco tramos de ingresos.

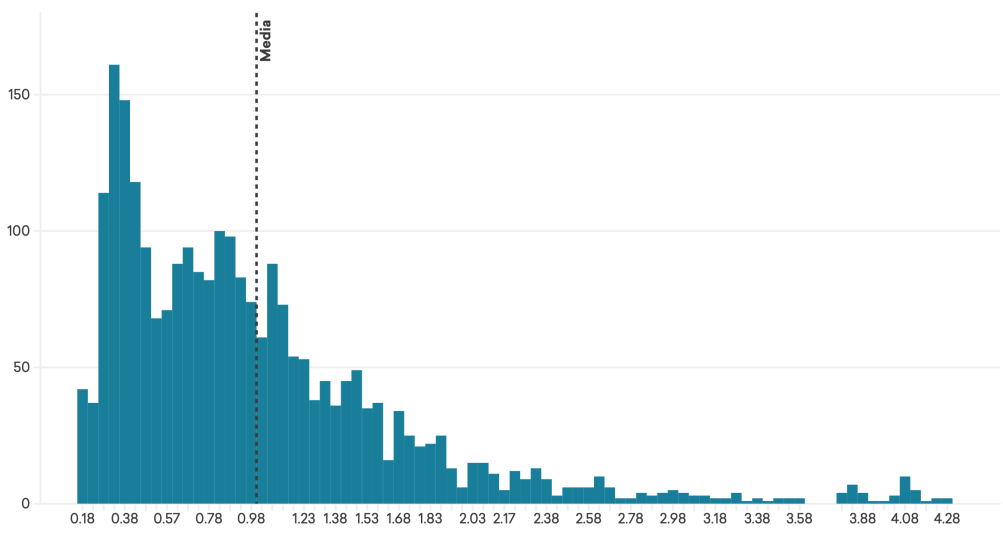
Antes de ponderar, aplicamos un filtro previo: eliminamos las entrevistas completadas en un tiempo excesivamente corto (normalmente, menos de un tercio de la duración mediana de la encuesta). Estas respuestas suelen ser de baja calidad porque el encuestado no ha leído las preguntas con atención. Además, complementamos las medidas de control con las que ya cuentan los paneles online con sistemas propios para detectar respuestas inconsistentes o de baja calidad.

Una vez aplicado el raking, controlamos que los pesos resultantes no sean extremos. Un peso demasiado alto significaría que una sola persona tiene una influencia desproporcionada sobre los resultados, lo cual introduce volatilidad y reduce la precisión efectiva de la muestra (Kish, 1992). Para evitarlo, limitamos el peso máximo a 4 veces la media y el peso mínimo a un cuarto de la media. Tras este recorte, se vuelve a aplicar el raking para restaurar la consistencia con los márgenes poblacionales.

Para evaluar la calidad de la ponderación utilizamos el efecto de diseño (DEFF), que cuantifica cuánta precisión se pierde por el hecho de usar pesos desiguales en lugar de una muestra equiprobable. Un DEFF de 1 significaría que los pesos no introducen ninguna pérdida; valores por encima de 2 o 3 empezarían a ser preocupantes. La Figura 3 muestra la distribución de los pesos finales de nuestra encuesta.

### Distribución de pesos de raking

DEFF: 1,52



Raking: género x edad, CCAA, tamaño municipio, ingresos, recuerdo.  
Cap: 4x. Rango: [0.187, 4.269]. n = 2.547.

**Figura 3.** Distribución de los pesos de raking. La línea discontinua marca la media (1.0). El DEFF de 1.52 indica una pérdida de precisión moderada y controlada.

## 3.1. Una estrategia en dos etapas

Hasta aquí hemos descrito la ponderación demográfica: corregir la muestra para que se parezca a España en términos de edad, género, territorio y renta. Pero hay una quinta dimensión que resulta fundamental para la estimación electoral: el recuerdo de voto.

Por ejemplo, si en nuestra muestra hubiera proporcionalmente más votantes del PSOE que los que realmente votaron al PSOE el 23-J de 2023, nuestra estimación estará sesgada a favor de ese partido (y lo mismo para cualquier otro). Cuando una encuesta no corrige este tipo de desequilibrio, los sesgos pueden ser persistentes y sustanciales, como ilustra el debate en torno a las estimaciones que se ha producido en España en los últimos años (Llaneras, 2025). Por eso las encuestas electorales suelen añadir el recuerdo de voto como una dimensión más de ponderación, usando como referencia los resultados oficiales de las últimas elecciones (Ministerio del Interior, 2023). Esta práctica no está exenta de debate (Durand et al., 2015; Escobar Mercado et al., 2014), pero es ampliamente utilizada por su capacidad de corregir la no respuesta diferencial por afinidad partidista (Rivero y Muñoz, 2022).

El problema es que no podemos hacerlo directamente. Una parte relevante de los encuestados no responde a la pregunta de recuerdo de voto: dice que no se acuerda o prefiere

no contestar. Si ponderamos con un recuerdo incompleto, los pesos se distorsionan. Pero para imputar ese recuerdo necesitamos un modelo que ya esté ponderado demográficamente.

Por eso utilizamos una estrategia en dos etapas. En la primera, ponderamos la muestra solo con las cuatro dimensiones demográficas. Estos pesos se utilizan para entrenar el modelo de imputación del recuerdo de voto. En la segunda, una vez completado el recuerdo de voto para todos los encuestados, añadimos una quinta dimensión de raking con los resultados oficiales del 23-J como referencia. Los pesos resultantes de esta segunda etapa son los que se usan en el resto de la estimación.

Esta separación evita un problema de circularidad: no se puede ponderar por recuerdo de voto antes de haberlo imputado, y no se puede imputar sin unas ponderaciones previas que corrijan los desequilibrios demográficos más básicos.

En la siguiente sección explicamos cómo abordamos precisamente ese problema: completar el recuerdo de voto para los encuestados que no lo declaran.

## 4. El ajuste del recuerdo de voto

---

En toda encuesta electoral hay una pregunta que parece sencilla pero esconde una enorme complejidad: «¿A qué partido votó usted en las últimas elecciones generales?». La respuesta a esta pregunta es nuestra única referencia verificable. A diferencia de la intención de voto, que mira al futuro y no puede contrastarse hasta que se celebren las elecciones, el recuerdo de voto se refiere a algo que ya ocurrió y cuyos resultados son públicos: los del 23 de julio de 2023. Eso lo convierte en un ancla fundamental del modelo, porque nos permite hacer dos cosas. Primero, usarlo como dimensión de ponderación para corregir el sesgo partidista de la muestra, como explicamos en la sección anterior. Y segundo, utilizarlo como uno de los predictores más potentes de la intención de voto actual: quién votó a un partido en 2023 tiene una probabilidad alta de volver a hacerlo.

Antes de preguntar por el partido, en nuestra encuesta preguntamos primero si el encuestado fue a votar o no. Este diseño en dos pasos no es casual. La literatura sobre diseño de cuestionarios ha mostrado que preguntar directamente por el partido sin filtrar antes por participación tiende a inflar la tasa de respuesta al recuerdo, porque muchos encuestados que en realidad no votaron se ven arrastrados a declarar un partido por presión social (Groves et al., 2009; Bernstein et al., 2001). Al separar las dos preguntas y ofrecer opciones que

no penalizan la abstención, se reduce significativamente la sobredeclaración de participación (Morin-Chassé et al., 2017). Esta separación nos permite, además, identificar con mayor precisión quiénes son los abstencionistas declarados, quiénes los votantes con recuerdo y quiénes los que participaron pero no quieren o no pueden recordar a quién votaron.

El problema es que, aun con este diseño, una parte significativa de los encuestados no contesta a la pregunta de recuerdo. Dice que no se acuerda, que prefiere no decirlo, o simplemente no responde. Esto no es ruido aleatorio. Las personas que han cambiado de partido, las que sienten que su voto anterior es socialmente poco deseable, o las que están menos implicadas políticamente tienden a ocultar o no declarar su recuerdo con más frecuencia (Durand et al., 2015; Urquizu, 2005). Además, la investigación sugiere que buena parte del desajuste entre el recuerdo declarado y los resultados reales se explica por participación diferencial en la encuesta: las personas más propensas a votar también son más propensas a responder encuestas (Gelman et al., 2016; Ansolabehere y Hersh, 2012; Selb y Munzert, 2013). En nuestros datos observamos que la tasa de no respuesta al recuerdo de voto varía de forma sistemática según la ideología, el nivel de interés político, la edad y otras variables observables. Esto apunta a que los datos faltantes no son aleatorios, sino que están relacionados con características medibles de los encuestados, lo que en terminología estadística se denomina *Missing at Random* (MAR) (Little y Rubin, 2019).

Además, incluso quienes sí contestan pueden no hacerlo con precisión. Existe una tendencia bien documentada a que el recuerdo de voto se distorsione con el tiempo: los encuestados tienden a declarar que votaron al partido que prefieren en el momento actual, o al que ganó las elecciones, un fenómeno conocido como reconciliación o sesgo de memoria (Durand et al., 2015; Pennay et al., 2023; Cabrera-Álvarez y Escobar, 2025). Todo esto significa que no podemos ni ignorar los datos faltantes ni asumir que las respuestas completas son perfectas.

¿Qué hacer, entonces, con los encuestados que no declaran su recuerdo? La opción más simple sería excluirlos del análisis y trabajar solo con quienes sí responden. Pero si la no respuesta no es aleatoria, descartar a estas personas sesga los resultados: equivale a asumir que se comportan igual que los respondientes, cuando sabemos que no es así. Otra alternativa sería repartir a los no respondientes proporcionalmente entre los partidos, según la distribución observada en quienes sí contestan. Pero esto también es problemático, porque asume que la composición partidista de los que no responden es idéntica a la de los que sí lo hacen, lo cual contradice la evidencia de que el perfil de ambos grupos difiere (Little y Rubin, 2019; Cabrera-Álvarez y Escobar, 2019). Necesitamos un enfoque que

tenga en cuenta las características de cada persona para estimar, de la forma más rigurosa posible, qué votó.

## 4.1. Cómo imputamos el recuerdo de voto

Nuestro enfoque consiste en utilizar toda la información disponible sobre cada encuestado para tratar de adivinar, con un alto nivel de certeza, a qué partido votó en 2023. Para ello empleamos un modelo de clasificación basado en *gradient boosted trees* (XGBoost), una técnica de aprendizaje automático que ha demostrado un buen rendimiento en problemas de clasificación con datos tabulares (Chen y Guestrin, 2016; Deng y Lumley, 2023).<sup>1</sup> Una ventaja práctica de XGBoost es que maneja de forma nativa los valores faltantes en las variables predictoras: cuando un encuestado no ha respondido a alguna pregunta (por ejemplo, su nivel de renta o su ideología), el algoritmo utiliza esa ausencia como señal informativa en lugar de descartar la observación.

De forma intuitiva, lo que hace el modelo es lo siguiente: observamos a todos los encuestados que sí declaran su recuerdo de voto, junto con el resto de sus respuestas. El modelo aprende qué combinaciones de estas variables se asocian con haber votado a cada partido. Después, aplica esas mismas asociaciones a los encuestados que no declararon su recuerdo para asignarles un voto probable.

Las variables que alimentan el modelo se seleccionan buscando un equilibrio entre poder predictivo y disponibilidad para todos los encuestados (incluidos los que no declaran recuerdo). Utilizamos tres bloques de información. El primero es el perfil sociodemográfico: edad, género, nivel educativo, Comunidad Autónoma, tamaño del municipio, renta del hogar, situación laboral, país de nacimiento, lengua habitual y religión. Estas variables capturan las bases estructurales del voto en España, donde la geografía, la edad y la clase social siguen siendo determinantes relevantes del comportamiento electoral (Groves et al., 2009). El segundo bloque recoge las actitudes políticas: la autoubicación ideológica en la escala izquierda-derecha, el interés por la política, la preferencia por un líder político y la opinión sobre el modelo territorial. Estas variables son especialmente útiles para discriminar entre partidos cercanos en el espacio ideológico. El tercer bloque, y probablemente el más informativo, son dos baterías de preguntas que se formulan a todos los encuestados:

---

<sup>1</sup>El uso de *gradient boosted trees* para la estimación electoral en España fue introducido de forma pionera por Gonzalo Rivero y Jordi Muñoz durante su etapa en el Centre d'Estudis d'Opinió de la Generalitat de Catalunya (2022–2024) (Rivero y Muñoz, 2022). Su modelo, que incluía la publicación abierta tanto de la metodología como del código completo de las estimaciones, ha sido una referencia e inspiración importante para el desarrollo del nuestro.

la probabilidad declarada de votar a cada uno de los principales partidos (en una escala de 0 a 3) y los vetos explícitos («a este partido no lo votaría nunca»). Estas baterías son particularmente valiosas porque permiten al modelo identificar el espacio de partidos plausibles para cada persona, incluso cuando no ha declarado su recuerdo.

Los encuestados que declararon voto en blanco o nulo se excluyen del modelo de imputación. Esta decisión no es solo práctica (su recuerdo ya es conocido), sino también metodológica. En la realidad, el voto en blanco y nulo representa una fracción muy pequeña del electorado, pero en las encuestas su peso relativo tiende a ser mayor (Ferrándiz y Camas García, 2019). Si se incluyeran como categorías en el modelo de clasificación, podrían funcionar como una especie de cajón de sastre: encuestados con perfiles ambiguos que no encajan claramente en ningún partido podrían acabar asignados a blanco o nulo con más frecuencia de la que les correspondería, distorsionando tanto esa categoría como las demás. Por eso los excluimos del modelo y los reincorporamos directamente con su valor original una vez completada la imputación del resto.

Una decisión de diseño importante es que nuestro modelo trata la abstención como una categoría más dentro del mismo problema de clasificación, en lugar de separar la predicción en dos pasos (primero si votó o no, y luego a quién). Esto responde a que la decisión de votar y la elección de partido no son independientes: hay personas que declaran haber votado pero en realidad se abstuvieron (por deseabilidad social), y modelar participación y voto de forma conjunta captura mejor estas dinámicas.

El modelo se entrena con los pesos demográficos calculados en la primera etapa de ponderación, lo que asegura que los patrones aprendidos reflejan a la población española y no solo a la composición bruta de la muestra.

Hay un aspecto del proceso de imputación que merece especial atención: cómo asignamos el voto a cada persona. El modelo no produce una respuesta única para cada encuestado, sino una distribución de probabilidades. Por ejemplo, para una persona concreta podría predecir un 45% de probabilidad de haber votado al PSOE, un 35% al PP, un 12% a Sumar y un 8% a Vox. La opción más simple sería asignar a esa persona el partido más probable (PSOE, en este caso). Pero esto tiene un problema: si todos los encuestados cuya primera opción es el PSOE reciben siempre esa asignación, se infla artificialmente el peso de los partidos mayoritarios y se subestima la variabilidad real del electorado.

Por eso optamos por una asignación estocástica: en lugar de elegir siempre el partido más probable, sorteamos el resultado según la distribución de probabilidades predicha.

La persona del ejemplo anterior sería asignada al PSOE el 45 % de las veces, al PP el 35 %, a Sumar el 12 % y a Vox el 8 %. Este mecanismo preserva la variabilidad natural del electorado y, además, es la base de la cuantificación de incertidumbre que empleamos más adelante: repetimos el sorteo 50 veces y medimos cuánto varían los resultados entre una realización y otra (Rubin, 1987; Buuren, 2018).

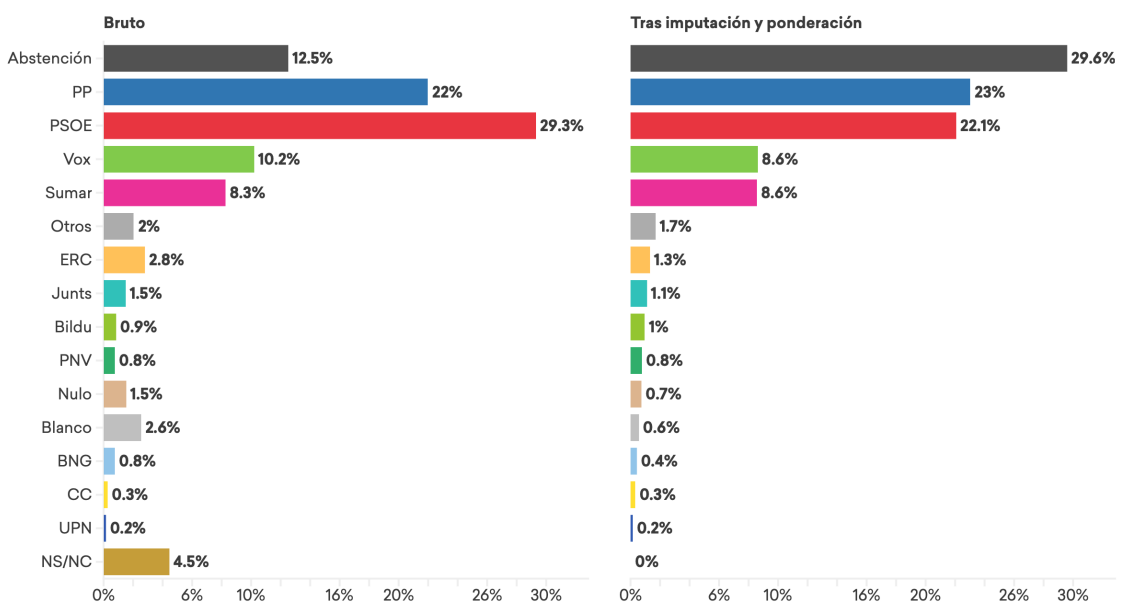
Para que el modelo tenga suficientes casos en cada categoría, agrupamos algunos partidos pequeños durante la fase de modelización: los partidos regionalistas se agrupan en una categoría conjunta y los partidos muy minoritarios en otra. Tras la imputación, desagregamos la categoría de partidos regionales según la distribución real de voto en cada Comunidad Autónoma, recuperando así el detalle territorial.

## 4.2. Validación y cierre del ciclo de ponderación

¿Cómo sabemos que la imputación es razonable? Antes de aplicar el modelo a los datos faltantes, evaluamos su capacidad predictiva mediante validación cruzada: dividimos los datos de entrenamiento (encuestados con recuerdo conocido) en cinco partes, entrenamos el modelo con cuatro de ellas y predecimos la quinta, rotando hasta cubrir todas. Esto nos permite medir la precisión del modelo sin hacer trampas, es decir, sin evaluar sobre los mismos datos con los que se entrenó. El modelo clasifica correctamente en torno al 76.9% de los encuestados en validación cruzada, con una precisión que varía por partido: los partidos más grandes tienden a predecirse con mayor tasa de acierto, mientras que los más pequeños o heterogéneos presentan tasas menores, lo cual es esperable dado el menor número de casos disponibles para el aprendizaje. La matriz de confusión completa, con el detalle de aciertos y errores por partido, se presenta en el Anexo A. El anexo incluye además diagnósticos complementarios: las curvas ROC (Figura 15) confirman que el modelo discrimina adecuadamente entre partidos y la curva de aprendizaje (Figura 17) muestra que la regularización es apropiada y no hay sobreajuste perjudicial. El modelo imputa el recuerdo de 110 encuestados (un 4.3 % del total). Pero más allá de esta validación interna, la forma más directa de comprobar la razonabilidad del resultado es observar cómo cambia la distribución del recuerdo de voto al aplicar el proceso completo de imputación y ponderación. La Figura 4 compara, para cada partido, su proporción en los datos brutos (panel izquierdo) y tras la imputación de NS/NC y la calibración por raking con los resultados oficiales del 23-J (panel derecho).

## Recuerdo de voto en 2023

Abril, 2026



% sobre población con derecho a voto.

**Figura 4.** Recuerdo de voto en 2023: comparación entre los datos brutos (izquierda) y los datos tras la imputación de NS/NC y la calibración por raking sobre los resultados oficiales del 23-J (derecha).

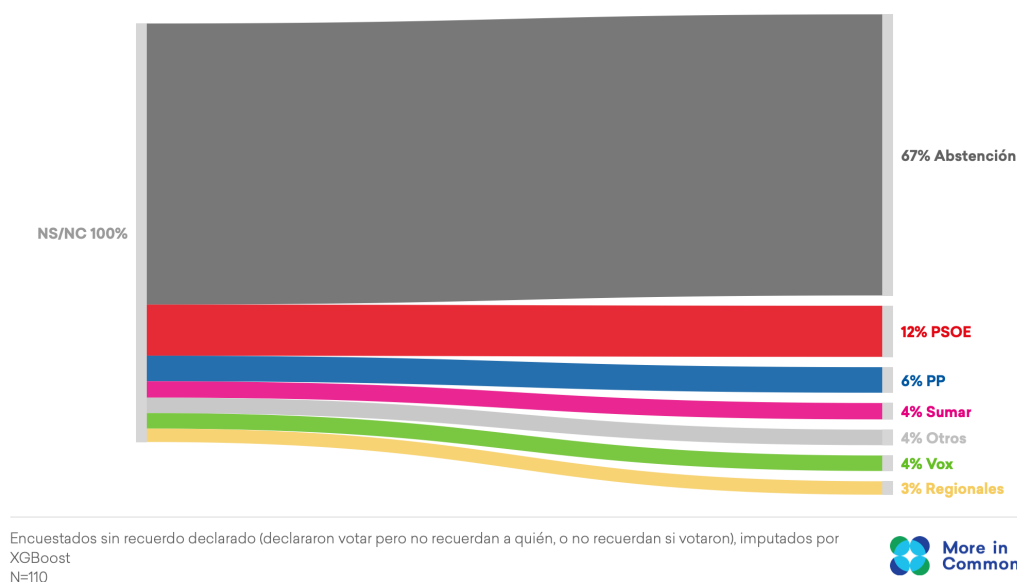
El panel izquierdo (Bruto) muestra el recuerdo tal como lo declaran los encuestados sin ningún ajuste. Ya en esta etapa se observan las distorsiones habituales: la abstención aparece infrarrepresentada (los encuestados tienden a declarar que votaron más de lo que realmente lo hicieron), mientras que algunos partidos aparecen sobrerrepresentados respecto a su resultado real y otros infrarepresentados. Una proporción visible de respondentes corresponde además a la categoría NS/NC. El panel derecho muestra el resultado tras todo el proceso: la imputación asigna un recuerdo a los encuestados que no lo declararon (la barra de NS/NC desaparece) y la calibración por raking, que añade el recuerdo como quinta dimensión de ponderación, fuerza a que los márgenes coincidan con los resultados oficiales del 23-J. La comparación entre ambos paneles ilustra la magnitud del ajuste que el modelo introduce en cada partido.

Lo relevante de esta progresión es que la calibración final no necesita aplicar correcciones extremas para alcanzar los márgenes oficiales, lo cual indica que las etapas previas (ponderación demográfica e imputación) están haciendo un trabajo razonable. Si la imputación estuviese muy desencaminada, la calibración tendría que forzar pesos desproporcionados, lo que se reflejaría en un DEFF elevado y en una pérdida importante de precisión. Los res-

sultados concretos pueden variar de una encuesta a otra, pero la lógica del proceso se mantiene.

La Figura 5 ofrece una perspectiva complementaria: muestra hacia dónde se dirigen los encuestados que no declararon su recuerdo de voto tras la imputación. El resultado más llamativo es que una proporción muy relevante son clasificados como abstencionistas. Esto es coherente con lo que sabemos sobre el perfil de los no respondientes: las personas menos implicadas políticamente son más propensas tanto a no votar como a no declarar su recuerdo en una encuesta. El resto se reparte entre los principales partidos, con PP y PSOE recibiendo las proporciones más altas, seguidos a distancia por partidos más pequeños.

### Imputación del recuerdo de voto: destino de los NS/NC



**Figura 5.** Destino de los encuestados sin recuerdo declarado (NS/NC) tras la imputación con XGBoost. La mayoría son clasificados como abstencionistas, seguidos por los dos principales partidos.

Con el recuerdo de voto completo para todos los encuestados, cerramos el ciclo que dejamos abierto en la sección anterior. Volvemos al proceso de ponderación y añadimos una quinta dimensión de raking: el recuerdo de voto, con los resultados oficiales del 23-J como referencia (Ministerio del Interior, 2023). Estos pesos finales, que incorporan tanto los equilibrios demográficos como los electorales, son los que se utilizan en todas las estimaciones posteriores. Conviene recordar que el objetivo de este proceso no es predecir el resultado de unas elecciones futuras, sino ofrecer la mejor fotografía posible de la situación actual del electorado, con sus incertidumbres explícitamente cuantificadas.

## 5. Imputando la intención de voto a los indecisos

---

Si la sección anterior miraba al pasado (qué votó cada persona en 2023), esta mira al presente: ¿a quién votaría hoy cada persona? La pregunta de intención de voto es el núcleo de toda encuesta electoral y, al mismo tiempo, la que más incertidumbre introduce. A diferencia del recuerdo de voto, que al menos tiene una respuesta verdadera verificable, la intención de voto se refiere a una decisión que aún no se ha tomado y que puede cambiar hasta el último momento.

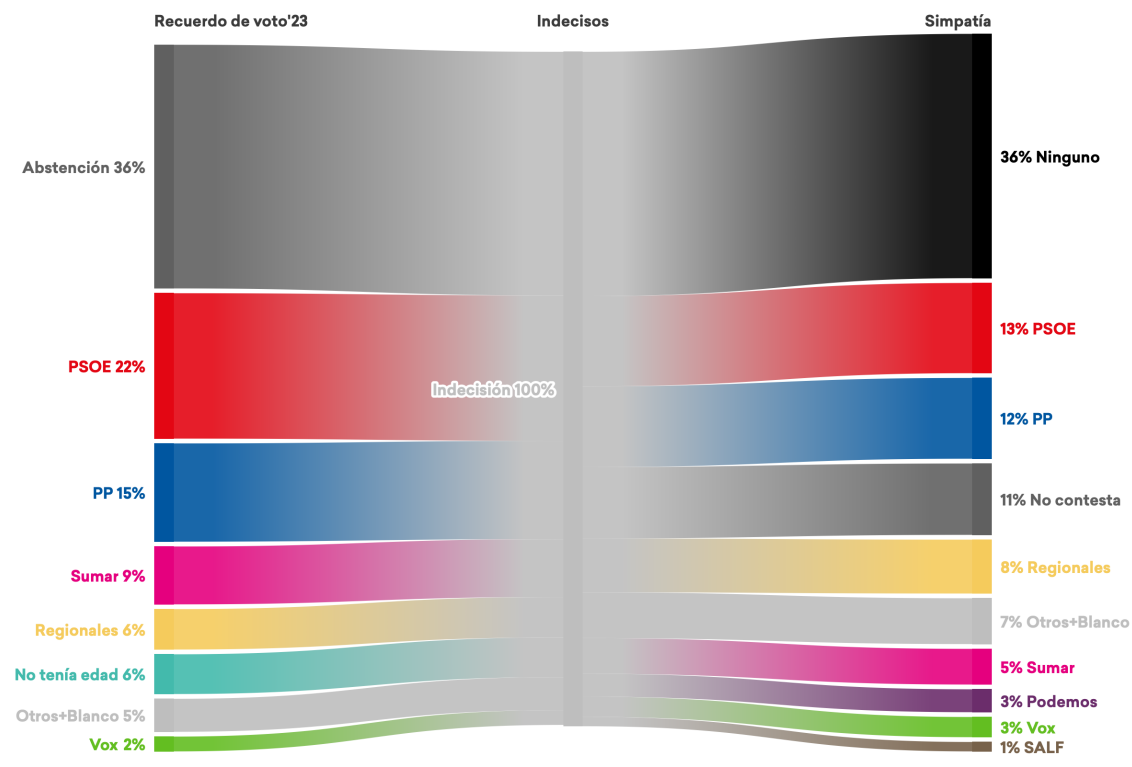
El problema central son los indecisos. Una proporción considerable de los encuestados responde «no lo sé» o «prefiero no contestar» cuando se les pregunta por su intención de voto (Ferrándiz y Camas García, 2019). Pero los indecisos no son un grupo homogéneo. En nuestro modelo distinguimos varios perfiles:

- Los que genuinamente no saben a quién votarían y responden NS/NC.
- Los que, al insistirles con una pregunta de simpatía («¿Con qué partido se siente más cercano?»), sí declaran una preferencia. Estos encuestados no estaban decididos pero sí tienen una inclinación, y la incorporamos como su intención de voto.
- Los que declaran que votarían en blanco o nulo, pero que en 2023 votaron a un partido concreto. Tratamos estos casos como indecisión encubierta: es más probable que estén dudando que que hayan transitado genuinamente hacia el voto en blanco.

La Figura 6 ilustra esta clasificación de forma visual. El gráfico muestra el recorrido de los encuestados que declaran no saber a quién votarían: a la izquierda, su recuerdo de voto en 2023; en el centro, su condición de indecisos; y a la derecha, el partido con el que simpatizan cuando se les formula la pregunta de seguimiento. Lo que revela es que los indecisos no son un grupo neutro o equidistante: tienen historias electorales y afinidades que los inclinan en direcciones concretas. Un porcentaje importante proviene de la abstención en 2023, lo que sugiere un perfil poco implicado políticamente. Entre quienes sí votaron, los principales partidos de origen suelen ser los más grandes. En la columna de simpatía, una parte significativa sigue sin declarar preferencia alguna (estos son los que el modelo necesita imputar), pero el resto se distribuye de forma desigual entre los partidos y se incorpora directamente con su respuesta de simpatía.

## ¿Qué votaron los indecisos en 2023? ¿Y con qué partido simpatizan?

Abril 2026



Encuestados que declaran no saber a quien votar en las próximas elecciones. N = 215

**Figura 6.** Recorrido de los indecisos: su recuerdo de voto en 2023 (izquierda), su condición de indecisos (centro) y su simpatía partidista declarada (derecha). Un porcentaje significativo proviene de la abstención y permanece sin preferencia clara.

### 5.1. Qué cambia respecto a la imputación del recuerdo

El enfoque general es el mismo que describimos en la sección anterior: un modelo XG-Boost que utiliza las características de cada encuestado para predecir su comportamiento electoral más probable, con asignación estocástica y las mismas baterías de probabilidad y veto como predictores clave (Chen y Guestrin, 2016; Hastie et al., 2001). Pero hay varias diferencias importantes.

La primera es que ahora disponemos de una variable predictora especialmente potente: el recuerdo de voto imputado en la sección anterior. Saber que alguien votó al PP en 2023 es una de las señales más fuertes para predecir su intención actual. Esta es precisamente la razón por la que el modelo se construye en etapas: la imputación del recuerdo alimenta la imputación de la intención, y no al revés.

La segunda diferencia está en las variables derivadas que construimos a partir de las ba-

terías de probabilidad de voto. Además de las probabilidades brutas para cada partido, calculamos indicadores sintéticos como la entropía de la distribución de probabilidades (Gill, 2005), que mide cuánto de dispersas o concentradas están las preferencias de un encuestado, la diferencia entre los dos partidos más probables (que refleja la claridad de la preferencia) y la coherencia entre la probabilidad máxima y el recuerdo de voto o el líder preferido. Estos indicadores ayudan al modelo a distinguir, por ejemplo, entre un indeciso que oscila entre dos partidos concretos y otro cuyas preferencias están completamente dispersas.

La tercera diferencia afecta a la agrupación de partidos. En el recuerdo de voto, Podemos formaba parte de la coalición de Sumar y se modelaban juntos. En la intención de voto actual, Podemos y SALF (Se Acabó La Fiesta) se tratan como categorías independientes, ya que son opciones diferenciadas en el panorama electoral presente.

La cuarta es un paso adicional que no existía en la imputación del recuerdo: la aplicación de restricciones a posteriori. Una vez que el modelo genera la distribución de probabilidades para cada encuestado, revisamos si es coherente con lo que esa persona ha declarado en otras partes de la encuesta. Si alguien ha dicho que nunca votaría a un partido, le asignamos una probabilidad de cero para ese partido y redistribuimos esa masa de probabilidad entre las opciones restantes. De este modo, el modelo no puede asignar votos a opciones que la propia persona ha descartado explícitamente.

Finalmente, aplicamos una corrección de sesgo que ajusta las probabilidades predichas utilizando los resultados oficiales del 23-J como referencia (Ministerio del Interior, 2023). Si el modelo tiende a sobreestimar un partido respecto a su resultado real en 2023, esta corrección reduce proporcionalmente esa tendencia (Shirani-Mehr et al., 2018). La corrección se aplica antes de la asignación estocástica, de modo que los sorteos reflejan ya las probabilidades calibradas.

Al igual que con el recuerdo de voto, este modelo se entrena con los pesos electorales completos (las cinco dimensiones de raking) y utiliza asignación estocástica con 50 re-sorteos para la cuantificación de incertidumbre (Rubin, 1987). En validación cruzada, el modelo de intención de voto alcanza una precisión global del 87.8%, considerablemente superior a la del modelo de recuerdo. Esta mejora se explica en buena parte por la incorporación del recuerdo de voto imputado como predictor. Las tasas de acierto por partido se detallan en el Anexo A, junto con las curvas ROC (Figura 16), que muestran una discriminación excelente, y la curva de aprendizaje del modelo (Figura 18). El modelo imputa la

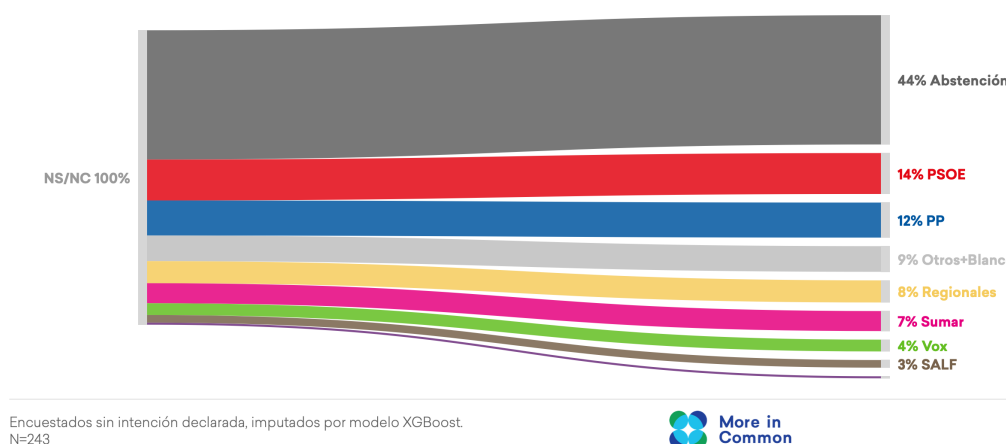
intención de 243 indecisos (9.5 % del total).

## 5.2. El destino de los indecisos tras la imputación

La Figura 7 muestra el destino final de los indecisos tras la imputación. La abstención sigue siendo el destino más frecuente, pero su peso es claramente menor que en el recuerdo de voto, mientras que los principales partidos reciben proporciones notables. Este reparto refleja que la intención de voto es un espacio más incierto que el recuerdo: los indecisos de hoy son personas que están genuinamente dudando, no simplemente personas que no quieren revelar una decisión ya tomada.

### Imputación de intención de voto: destino de los NS/NC

Abril, 2026

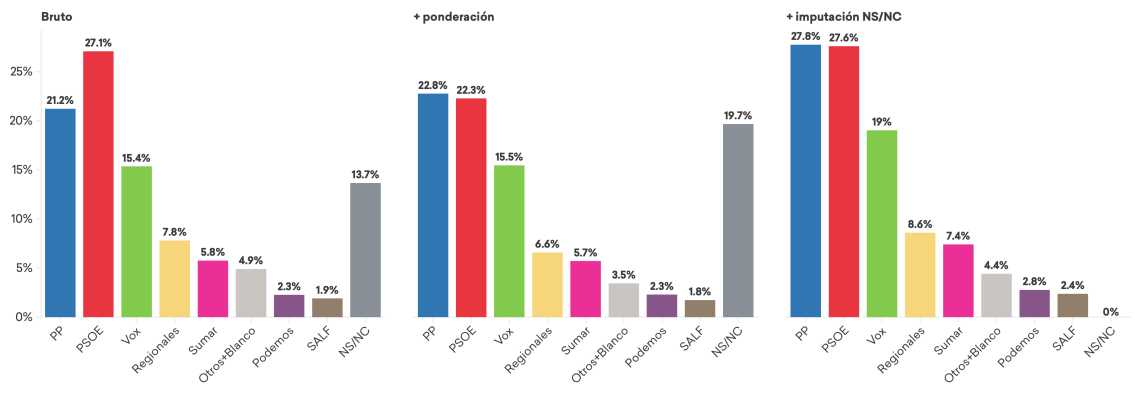


**Figura 7.** Destino de los encuestados sin intención de voto declarada (NS/NC) tras la imputación con XGBoost. La distribución es más dispersa que en el recuerdo de voto, reflejando la mayor incertidumbre del presente.

La Figura 8 ofrece una visión de conjunto: la intención de voto por partido en tres etapas sucesivas (datos brutos, tras la ponderación por raking y tras la imputación de los NS/NC). La barra de NS/NC, que en los datos brutos representa una proporción considerable, desaparece completamente tras la imputación, redistribuida entre los partidos. Lo relevante es observar cómo cambian las proporciones de cada partido al pasar de los datos brutos a los datos completos: algunos partidos ganan proporcionalmente con la imputación (porque los indecisos se les asignan con mayor frecuencia) y otros pierden. Estas diferencias son precisamente el valor añadido de la imputación frente a limitarse a ignorar a los indecisos o repartirlos proporcionalmente.

## Intención de voto

Abril, 2026



% sobre población con derecho a voto.

**Figura 8.** Intención de voto: comparación entre datos brutos, ponderados y tras la imputación de los indecisos. La categoría NS/NC desaparece completamente tras la imputación.

Con la intención de voto completa para todos los encuestados, el siguiente paso es responder a una pregunta distinta pero igualmente crucial: de todas estas personas que declaran (o se les asigna) una intención de voto, ¿cuántas irán realmente a votar?

## 6. ¿Quién irá a votar?

Hasta este punto disponemos de una intención de voto completa para cada encuestado: quienes declararon su preferencia, quienes fueron recuperados por la pregunta de simpatía y quienes fueron imputados por el modelo. Pero saber a quién votaría cada persona no es suficiente. La pregunta que falta es quién irá realmente a las urnas.

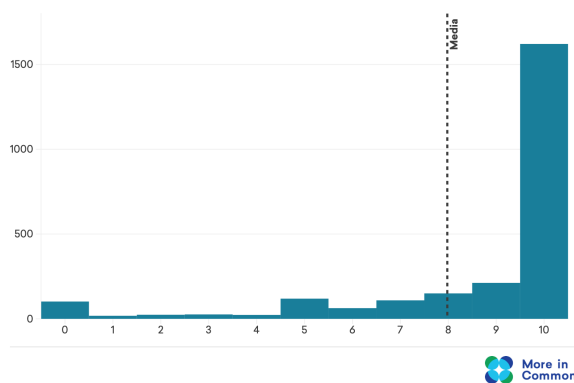
Esta es, posiblemente, la pieza más difícil de todo el proceso de estimación. La participación electoral es intrínsecamente difícil de predecir porque depende de factores que una encuesta capta solo parcialmente: la motivación del momento, la percepción de utilidad del voto, las condiciones meteorológicas del día de las elecciones, la intensidad de la campaña en la recta final... Y, sobre todo, porque la autodeclaración de participación en encuestas es notoriamente poco fiable. Las personas tienden a sobredeclarar que irán a votar, en parte por deseabilidad social (votar se percibe como un deber cívico (Blais y Galais, 2016)) y en parte por sobreestimación sincera de sus propias intenciones (Bernstein et al., 2001; Holbrook y Krosnick, 2010; Karp y Brockington, 2005). Este fenómeno no es exclusivo de ningún país: estudios con datos de 128 encuestas postelectorales en 43 países muestran que la tasa de participación declarada excede casi siempre la tasa oficial (Selb

y Munzert, 2013).

La Figura 9 ilustra este problema con claridad. El panel izquierdo muestra la distribución de la autopercepción de participación declarada por los encuestados (en una escala de 0 a 10): la inmensa mayoría se concentra en los valores más altos, con un pico enorme en 10 («seguro que votaré»). Si tomásemos estas respuestas al pie de la letra, la participación estimada superaría con creces cualquier registro histórico. El panel derecho muestra la distribución de la probabilidad de participación estimada por nuestro modelo (XGBoost): aunque la masa sigue concentrada en valores altos, el modelo dispersa el pico de la auto-declaración en una distribución continua, reflejando que no todos los que declaran certeza absoluta acabarán votando con la misma probabilidad. Las líneas discontinuas marcan la media histórica de participación y la media muestral, permitiendo contextualizar ambas distribuciones.

#### Auto-percepción de participar en las próximas elecciones

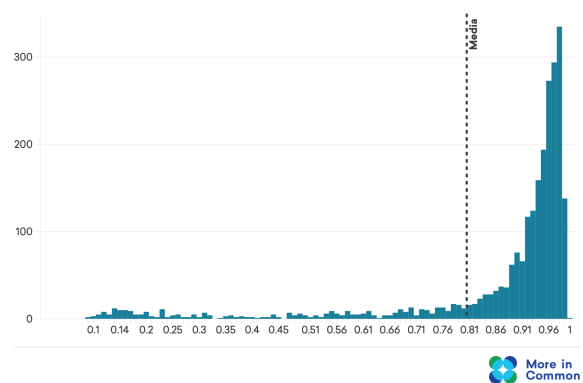
Probabilidad declarada por el encuestado



(a) Autopercepción declarada (escala 0–10).

#### Estimación XGBoost de participar en las próximas elecciones

Probabilidad calibrada del modelo



(b) Estimación del modelo XGBoost (probabilidad continua).

**Figura 9.** Distribución de la probabilidad de participación: autopercepción declarada por los encuestados (izquierda) frente a la estimación del modelo XGBoost (derecha). La autodeclaración se concentra en los valores más altos, mientras que el modelo produce una distribución más realista.

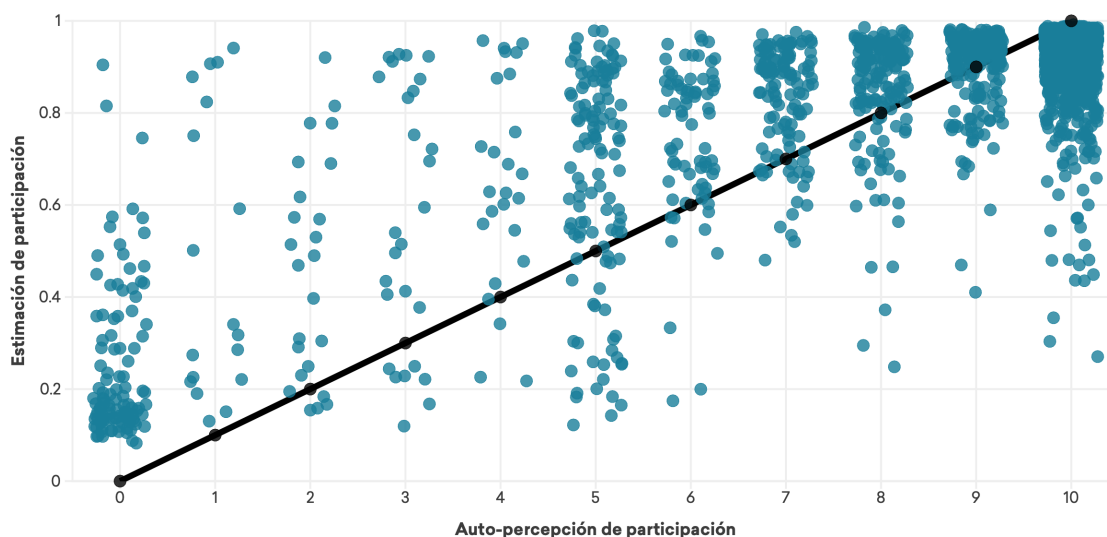
Para modelar la probabilidad individual de votar utilizamos, al igual que en las secciones anteriores, un modelo XGBoost (Chen y Guestrin, 2016). En este caso el modelo funciona como una regresión (no como un clasificador): a partir de las características de cada encuestado, predice un valor continuo entre 0 y 1 que representa su probabilidad estimada de acudir a votar. La curva de aprendizaje del modelo (Figura 19 en el Anexo) confirma una convergencia rápida y una brecha reducida entre entrenamiento y validación, indicando un modelo bien regularizado.

La Figura 10 muestra cómo actúa esta corrección persona a persona. Cada punto es un en-

cuestado: el eje horizontal representa lo que declara (su autopercepción de participación, en la escala discreta de 0 a 10) y el eje vertical lo que el modelo estima. La línea escalonada discontinua marca el escenario sin corrección, donde la predicción del modelo coincidiría exactamente con lo que dice el encuestado (convertido a proporción). Los puntos que quedan por debajo de esa línea son personas a quienes el modelo asigna una probabilidad de votar menor de lo que ellas mismas declaran: el modelo detecta, a partir de sus otras respuestas, que su participación real probablemente será inferior a la que expresan. Es el caso de la mayoría de quienes dicen 10 («seguro que votaré»): el modelo rebaja su probabilidad porque no todos los que dicen eso acabarán votando. En cambio, los puntos por encima de la línea corresponden a personas cuyo perfil sugiere una participación mayor de la que declaran. El resultado es una distribución corregida que aprovecha toda la información disponible sobre cada persona, no solo su respuesta a una única pregunta.

### Corrección sobre la auto-percepción de participación

La línea diagonal sería la probabilidad sin corrección.



Cada punto es un encuestado.

**Figura 10.** Corrección del modelo XGBoost sobre la autopercepción de participación. Cada punto es un encuestado. La línea diagonal representa la ausencia de corrección (predicción = autodeclaración). Los puntos por debajo de la diagonal indican que el modelo estima una participación menor a la declarada; los puntos por encima, mayor.

## 6.1. Qué información utiliza el modelo

Las variables que alimentan el modelo de participación combinan señales de distinta naturaleza. La más potente es el comportamiento pasado: si una persona votó en 2023, tiene una probabilidad alta de volver a hacerlo. La intención de voto declarada (o imputada)

también es muy informativa: quienes se declaran abstencionistas o a quienes el modelo de intención asignó la categoría de abstención reciben, lógicamente, probabilidades de participación más bajas. Este indicador por sí solo explica una fracción muy relevante de la capacidad predictiva del modelo.

A estas señales individuales se añade un predictor ecológico: la participación real registrada en cada Comunidad Autónoma en las elecciones del 23-J. Este dato captura diferencias regionales en cultura política que las variables individuales no recogen del todo (la participación en Cantabria o La Rioja históricamente supera el 75 %, mientras que en Ceuta y Melilla rara vez alcanza el 55 %). El resto de predictores incluye las variables demográficas habituales (edad, género, nivel educativo, renta, tamaño del municipio), las actitudes políticas (ideología, interés por la política, norma de participación) y las baterías de probabilidad y veto por partido.

Una decisión importante es que no calibramos la participación estimada para que coincida con un nivel oficial determinado. El modelo captura la propensión relativa de cada persona a votar (quién es más o menos probable que vote respecto a los demás), no un pronóstico del nivel agregado de participación. Esto es coherente con la filosofía general del modelo: ofrecer una fotografía del presente, no una predicción sobre un día de elecciones que aún no ha llegado.

## 6.2. De la probabilidad al voto: cómo decidimos quién votará

Hasta aquí, cada encuestado tiene asignada una probabilidad continua de votar. Pero para calcular las cuotas de voto necesitamos dar un paso más: decidir quién vota y quién no.

La forma más habitual de hacerlo en *"la industria de la demoscopia"* es utilizar escalas tipo Perry-Gallup, que combinan varias preguntas (intención de votar, interés político, comportamiento pasado) en una puntuación discreta, y luego aplican un umbral fijo para clasificar a los encuestados como «votantes probables» (Pew Research Center, 2016). El problema de este enfoque es que el umbral lo elige el analista, se aplica por igual a todos y no tiene en cuenta que el punto de corte adecuado depende del historial de cada persona.

Pongamos un ejemplo concreto para entender nuestro enfoque alternativo. Imaginemos dos encuestados a quienes el modelo les ha asignado la misma probabilidad de votar: un 75 %. El primero votó en 2023; la segunda se abstuvo. ¿Deberíamos tratarlos igual? Probablemente no. Un 75 % en alguien que ya votó antes no es especialmente sorprendente: la mayoría de los votantes anteriores muestran probabilidades altas, así que un 75 % es un

valor más bien modesto para ese grupo. Pero un 75% en alguien que se abstuvo en 2023 es una señal fuerte: pocas personas de ese grupo alcanzan probabilidades tan altas. Dicho de otra forma, la misma probabilidad tiene un significado distinto según el grupo al que pertenece la persona.

Nuestro enfoque formaliza exactamente esta intuición. En lugar de elegir un umbral fijo, dejamos que los datos nos digan dónde ponerlo, y lo hacemos de forma diferenciada para cada grupo. El procedimiento se desarrolla en tres pasos.

**Paso 1: mirar cómo se distribuyen las probabilidades en cada grupo.** Tenemos encuestados cuyo comportamiento declarado en 2023 conocemos (con las limitaciones propias de cualquier autodeclaración): los que dicen haber votado, los que dicen haberse abstenido y los que no tenían edad. Para cada uno de estos tres grupos, observamos cómo se distribuyen las probabilidades de participación que el modelo les ha asignado. El panel superior de la Figura 11 muestra estas distribuciones: la curva de los votantes 2023 se concentra en la zona alta (hacia la derecha), mientras que la curva de los abstencionistas 2023 se extiende más hacia la zona baja. Las curvas se solapan en la zona intermedia: ahí es donde está la incertidumbre real, y donde necesitamos un criterio para decidir.

**Paso 2: comparar las curvas.** Para cada valor de probabilidad, nos preguntamos: ¿cuántas veces más frecuente es este valor entre los votantes que entre los abstencionistas? Este cociente se denomina *likelihood ratio* (cociente de verosimilitud). Si en un punto concreto la curva de los votantes está muy por encima de la de los abstencionistas, el cociente será alto: una persona con esa probabilidad se parece mucho más a un votante. Si las dos curvas están a la misma altura, el cociente será cercano a 1: no podemos distinguir. Y si la curva de los abstencionistas domina, el cociente será bajo: esa probabilidad es más propia de un abstencionista. El panel inferior de la Figura 11 muestra esta curva de comparación: crece a medida que la probabilidad aumenta, porque las probabilidades altas son cada vez más frecuentes entre votantes y menos entre abstencionistas. En términos formales:

$$LR(\hat{p}) = \frac{f_{\text{votantes}}(\hat{p})}{f_{\text{abstencionistas}}(\hat{p})}$$

**Paso 3: aplicar un nivel de exigencia distinto a cada grupo.** El cociente de verosimilitud nos dice cuánto se parece alguien a un votante frente a un abstencionista. Pero falta decidir cuánta evidencia necesitamos para clasificarle como votante. Y aquí es donde entra el historial del grupo. Para los votantes de 2023, la base de partida es alta: la mayoría dicen que volverán a votar, así que no necesitamos mucha evidencia adicional. El listón es bajo.

Para los abstencionistas de 2023, la base es baja: muy pocos declaran certeza de votar, así que necesitamos evidencia muy fuerte para creer que esta vez sí irán. El listón es alto.

Calculamos este nivel de exigencia empíricamente a partir de los propios datos: miramos qué proporción de cada grupo declara la máxima certeza de participar ( $\pi$ ). El umbral para clasificar como votante queda entonces:

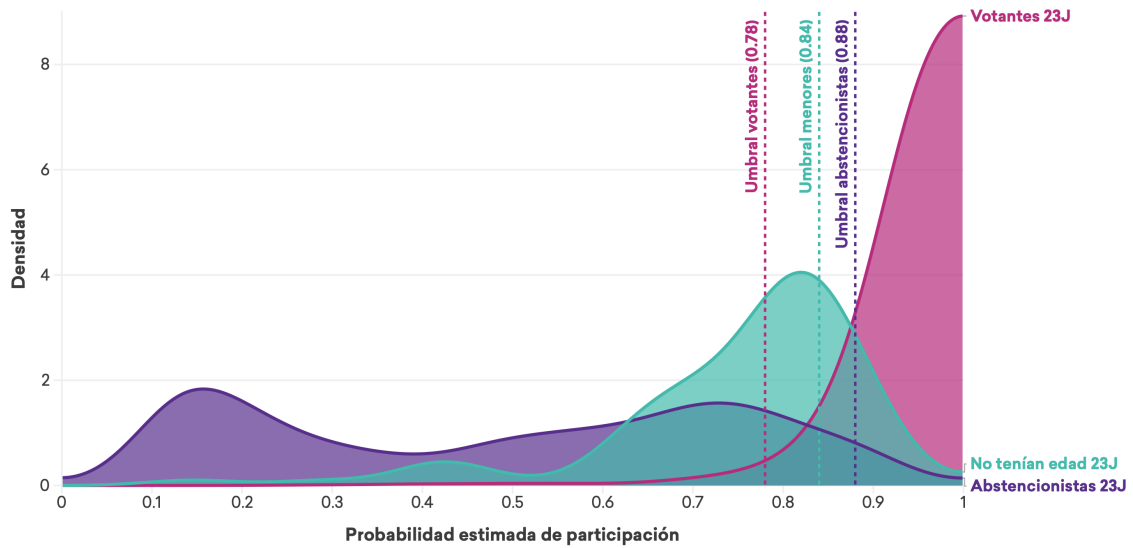
$$LR(\hat{p}) > \frac{1 - \pi}{\pi} \Rightarrow \text{votante}$$

Cuando  $\pi$  es alto (como en los votantes de 2023), la fracción  $(1 - \pi)/\pi$  es pequeña y el umbral es fácil de superar. Cuando  $\pi$  es bajo (como en los abstencionistas), la fracción es grande y el umbral exige mucha más evidencia. Los jóvenes que no tenían edad en 2023 se sitúan en un punto intermedio.

Lo importante de este enfoque es que ninguno de los números es elegido por el analista. Las distribuciones se estiman a partir de los datos del modelo, las probabilidades previas se calculan a partir de las respuestas de los encuestados, y los umbrales se derivan de la regla de decisión. Todo proviene de los datos. Esto no significa que el método esté completamente libre de decisiones: la definición del subconjunto de alta certeza para calcular las probabilidades previas, o la propia decisión de usar tres grupos en lugar de uno, son opciones por las que hemos optado tras valorar varias alternativas. Pero son opciones técnicas con justificación estadística, no juicios de valor sobre cuánta gente «debería» votar.

Con esta clasificación, cada encuestado queda etiquetado como votante o abstencionista. En la siguiente sección, combinamos esta etiqueta con la intención de voto y los pesos de la encuesta para producir la estimación final de cuotas de voto.

## Distribución de la probabilidad estimada según comportamiento en 2023

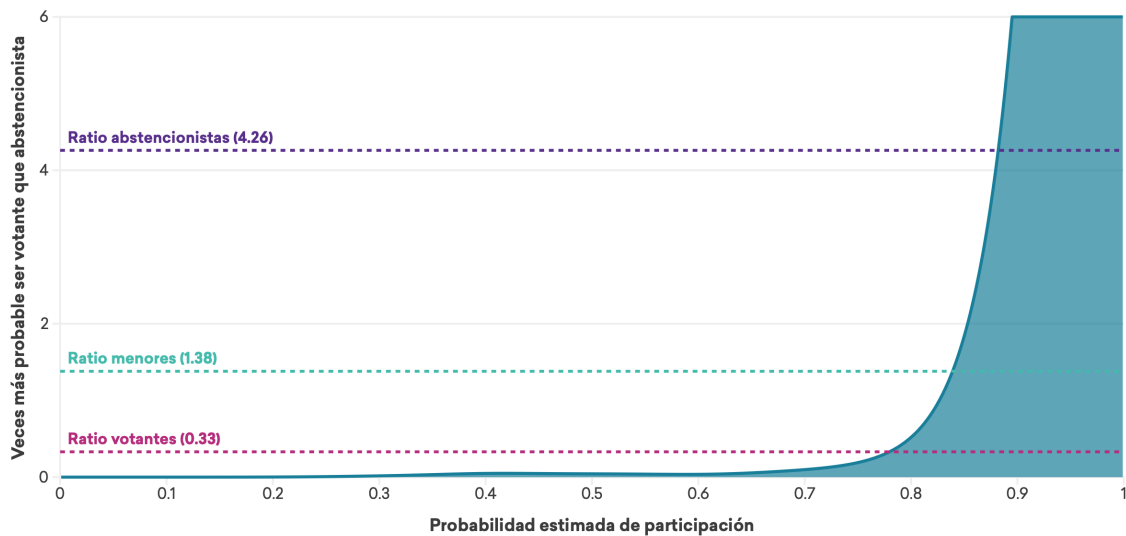


Curvas de densidad de la probabilidad estimada de participación en cada grupo. Líneas verticales: umbrales del escenario principal.

(a) Distribuciones de densidad de la probabilidad de participación para los tres grupos, con sus umbrales diferenciados marcados por líneas verticales.

## Razón de verosimilitud (LR) por probabilidad estimada

Cruce de la curva LR con el ratio de odds de cada grupo



La curva mide cuántas veces más probable es observar esa probabilidad estimada en un votante que en un abstencionista. El umbral de cada grupo se define donde la curva LR cruza su ratio de odds.

(b) Curva de cociente de verosimilitud y puntos de corte derivados de las probabilidades previas empíricas de cada grupo.

**Figura 11.** Clasificación de participación por cociente de verosimilitud (*likelihood ratio*). Panel superior: distribuciones de densidad de la probabilidad de participación para los tres grupos (votantes 2023, no tenían edad y abstencionistas 2023), con sus umbrales diferenciados marcados por líneas verticales. Panel inferior: curva de cociente de verosimilitud (cuántas veces más probable es ser votante que abstencionista para cada nivel de probabilidad) y puntos de corte derivados de las probabilidades previas empíricas de cada grupo, indicados con su valor de  $\pi$  y su ratio correspondiente.

## 7. La estimación final

---

Llegados a este punto, disponemos de toda la información necesaria para producir una estimación de voto. Para cada encuestado tenemos: una intención de voto completa (declarada o imputada), una probabilidad individual de participación (estimada por el modelo de XGBoost) y una clasificación como votante o abstencionista (derivada de los umbrales de verosimilitud). Todo ello con unos pesos de encuesta que corrigen los desequilibrios demográficos y electorales de la muestra. En esta sección explicamos cómo ensamblamos estas piezas para obtener las cuotas de voto finales y, sobre todo, cómo cuantificamos la incertidumbre que rodea a esas cifras.

### 7.1. Cómo calculamos la estimación de voto

El cálculo de la cuota de voto se basa en un concepto sencillo: el peso efectivo de cada encuestado. Este peso combina dos elementos: el peso de la encuesta (que corrige los desequilibrios de la muestra, como vimos en la sección 3) y la probabilidad de participación (que refleja cuánto de probable es que esa persona realmente vote). En términos formales:

$$w_{\text{eff},i} = w_{\text{raking},i} \times \hat{p}_{\text{turnout},i}$$

La lógica es intuitiva: una persona que es muy probable que vote y que además está infrarrepresentada en la muestra contribuye más al resultado que una persona que probablemente no votará y que está sobrerrepresentada. El peso efectivo captura ambas dimensiones simultáneamente.

Una vez calculado el peso efectivo, distinguimos dos tipos de encuestados según cómo contribuyen a la cuota de cada partido:

- Los encuestados que declararon directamente su intención de voto aportan el 100 % de su peso efectivo al partido que indicaron. Si alguien dice que votará al PP, todo su peso va al PP.
- Los encuestados cuya intención fue imputada por el modelo no reciben una asignación fija, sino que su peso se reparte entre los partidos en proporción a las probabilidades predichas. Si el modelo estimó un 45 % de probabilidad para el PSOE y un 35 % para el PP, esos porcentajes del peso efectivo van a cada partido respectiva-

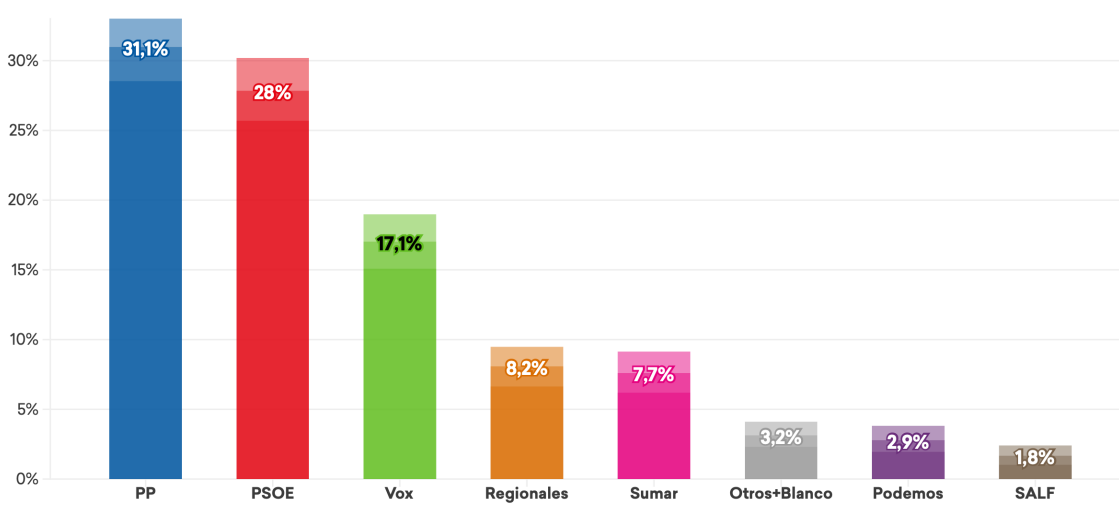
mente. Esto es coherente con la asignación estocástica que describimos en secciones anteriores: tratamos a los encuestados imputados como personas cuyo voto es genuinamente incierto, no como votantes de un único partido.

La cuota de voto de cada partido se expresa como porcentaje del voto válido total: la suma de los pesos efectivos asignados a ese partido dividida entre la suma de los pesos asignados a todos los partidos (excluyendo el voto nulo, que se reporta por separado).

La Figura 12 muestra el resultado principal de la estimación. Para cada partido se representan tres puntos verticalmente: el punto central marca la estimación puntual y los puntos extremos delimitan el intervalo de confianza al 95 %. Este intervalo se calcula mediante las reglas de Rubin, que describimos en la siguiente subsección.

### Estimación de voto

Participación: 68,3% [65,8-72,12%]



% sobre voto válido.

Intervalos de confianza al 95%.

La incertidumbre se ha calculado según las reglas de Rubin para imputación múltiple (M = 50). Trabajo de campo del 21 al 25 de abril de 2026. N=2547

**Figura 12.** Estimación de voto (escenario principal). Para cada partido se representan tres puntos: el central indica la estimación puntual y los extremos los límites del intervalo de confianza al 95 % derivado de las reglas de Rubin (M=50). Las cifras se expresan como porcentaje del voto válido.

Algunos aspectos que merece la pena destacar del gráfico más arriba. La estimación de participación se indica en el subtítulo del gráfico, con su propio intervalo de confianza. Las cuotas se expresan sobre voto válido, de modo que suman 100 % entre todos los partidos. La distancia vertical entre los puntos extremos varía por partido: los partidos con más encuestados y menos imputación tienden a tener intervalos más estrechos, mientras que los partidos más pequeños o con mayor proporción de voto imputado muestran más incerti-

dumbre.

## 7.2. Cuantificación de la incertidumbre

Una estimación sin intervalo de confianza es una estimación incompleta. Decir que un partido obtendría un 28% no significa lo mismo si el intervalo es 27-29% que si es 24-32%. Los intervalos de confianza son los que permiten al lector distinguir entre diferencias reales y diferencias que podrían deberse al azar o a la incertidumbre del modelo.

En nuestro caso, la incertidumbre tiene dos orígenes distintos. El primero es el error muestral: incluso con un modelo perfecto, una muestra diferente de la población habría dado resultados ligeramente distintos. Este es el tipo de error que recogen los márgenes de error habituales en las encuestas. El segundo es el error de imputación: dado que hemos estimado el recuerdo de voto y la intención de voto para una parte de los encuestados mediante asignación estocástica, un sorteo diferente habría producido asignaciones ligeramente distintas y, por tanto, cuotas ligeramente diferentes.

Para combinar ambas fuentes de incertidumbre utilizamos las reglas de Rubin para imputación múltiple (Rubin, 1987; Buuren, 2018). El procedimiento es conceptualmente sencillo: repetimos la estimación completa 50 veces ( $M = 50$ ). En cada repetición, volvemos a sortear el recuerdo de voto (a partir de las probabilidades del modelo de recuerdo), volvemos a sortear la intención de voto (a partir de las probabilidades del modelo de intención), volvemos a clasificar a cada encuestado como votante o abstencionista, y recalculamos las cuotas de voto. Cada repetición produce un resultado ligeramente distinto, porque los sorteos son aleatorios.

La estimación puntual final es simplemente la media de los 50 resultados. Si las cuotas apenas varían entre repeticiones, la imputación introduce poca incertidumbre y los intervalos serán estrechos. Si varían mucho, los intervalos serán más anchos, y eso nos dice que la estimación depende de forma sensible de cómo se resolvieron los datos faltantes.

La incertidumbre total se calcula combinando dos componentes: la varianza dentro de cada repetición ( $\bar{U}$ , que captura el error muestral y existiría aunque la imputación fuese perfecta) y la varianza entre repeticiones ( $B$ , que captura cuánto cambian los resultados según qué sorteo se realizó). La varianza total y el intervalo de confianza se construyen así:

$$T = \bar{U} + \left(1 + \frac{1}{M}\right) B$$

$$IC_{95\%} = \bar{Q} \pm 1,96 \times \sqrt{T}$$

donde  $\bar{Q}$  es la media de las 50 estimaciones puntuales. En definitiva, el intervalo de confianza refleja tanto la incertidumbre inherente al muestreo como la que introduce el proceso de imputación.

La transparencia sobre la incertidumbre es, en nuestra opinión, tan importante como la propia estimación. Una encuesta que presenta un único número sin márgenes transmite una falsa impresión de precisión. Al mostrar los intervalos de confianza, invitamos al lector a interpretar los resultados como lo que son: la mejor aproximación posible al estado actual de las preferencias del electorado, con márgenes honestos que reflejan tanto lo que sabemos como lo que no.

## 8. Conclusión

---

A lo largo de este informe hemos descrito, paso a paso, el modelo de estimación electoral que hemos desarrollado en More in Common España: desde la ponderación de la muestra hasta la clasificación de votantes y abstencionistas, pasando por la imputación del recuerdo de voto y de la intención de los indecisos. En cada etapa hemos intentado hacer dos cosas: explicar qué hacemos y por qué lo hacemos así, incluyendo las alternativas que descartamos y las razones para descartarlas.

Tres principios han guiado el diseño del modelo. El primero es la transparencia: publicar no solo los resultados, sino el método completo, con sus supuestos y sus limitaciones, para que cualquiera pueda evaluarlo y, si lo considera oportuno, cuestionarlo. El segundo es dejar que los datos hablen: en cada punto donde el modelo necesita una decisión, hemos buscado un criterio empírico en lugar de imponer uno arbitrario.; los umbrales de participación derivados por cociente de verosimilitud, la asignación estocástica en lugar de la determinista, o la calibración contra resultados oficiales conocidos son ejemplos de este principio. El tercero es la honestidad sobre la incertidumbre: presentar intervalos de confianza en lugar de un único número que transmita una falsa impresión de precisión.

Dicho esto, es importante ser claros sobre lo que el modelo no puede hacer y dónde sus limitaciones son más relevantes.

La muestra de nuestra encuesta procede de un panel online, lo cual implica que hay perfiles de la población que están sistemáticamente infrarrepresentados o ausentes. La ponderación corrige estos desequilibrios en la medida de lo posible, pero no puede compensar completamente los sesgos que introduce el modo de administración que, por otra parte, existirían también si utilizáramos otro modo. Los resultados del modelo dependen de las especificaciones elegidas: los hiperparámetros, las variables seleccionadas y la forma en que se agrupan los partidos influyen en los resultados. Hemos buscado configuraciones robustas mediante validación cruzada, pero otras elecciones razonables podrían producir estimaciones ligeramente distintas. La participación sigue siendo la mayor fuente de incertidumbre. El recuerdo de voto, que es un factor desde el que se calibra el modelo, se basa en lo que los encuestados declaran, no en registros verificados. Las distorsiones de memoria y la deseabilidad social introducen ruido que la imputación atenúa pero no elimina. Los partidos nuevos o de reciente aparición carecen de un historial electoral contra el que calibrar las predicciones, lo que aumenta la incertidumbre de sus estimaciones. Y, sobre todo, y como no nos hemos cansado de repetir a lo largo de todo el documento, una estimación electoral no es una predicción: refleja el estado de las preferencias en el momento de la encuesta, no lo que ocurrirá el día de las elecciones.

Este modelo no es un ejercicio cerrado. Es una herramienta viva, diseñada para mejorar con cada nueva ola de datos. Las decisiones que hemos tomado hoy pueden revisarse mañana a la luz de nueva evidencia, de mejores técnicas o de críticas fundamentadas. Publicamos este informe precisamente para facilitar ese proceso: si alguien identifica un punto débil, una alternativa mejor o un error en nuestro razonamiento, queremos saberlo. La mejor forma de hacernos llegar cualquier comentario, crítica o sugerencia es enviando un email a [hola@moreincommon.com](mailto:hola@moreincommon.com).

Durante mucho tiempo, la investigación por encuesta de calidad ha sido un recurso reservado a partidos políticos, grandes medios y empresas con presupuestos elevados. La sociedad civil, las organizaciones sin ánimo de lucro, los investigadores independientes y los propios ciudadanos han quedado fuera de un debate que, sin embargo, les afecta directamente. Democratizar la demoscopia significa no solo publicar resultados, sino compartir los datos y explicar con detalle los métodos para que cualquiera pueda entenderlos, cuestionarlos y construir sobre ellos. Significa que la opinión pública deje de ser una caja negra y se convierta en un bien común.

Este informe, junto con la publicación de nuestros microdatos, es nuestra (primera) contribución a ese objetivo. No es un ejercicio cerrado: es una invitación abierta a que otros lo examinen, lo mejoren y lo utilicen.

## A. Validación de los modelos de clasificación

---

En esta sección se presentan los resultados detallados de la validación cruzada (5 iteraciones) de los modelos XGBoost utilizados para la imputación del recuerdo de voto y de la intención de voto. Las matrices de confusión muestran, para cada partido, qué proporción de encuestados con respuesta conocida son clasificados correctamente por el modelo cuando se evalúan sobre datos que no formaron parte del entrenamiento. Cada fila representa el valor real declarado por el encuestado y cada columna la predicción del modelo. Los valores de la diagonal principal indican clasificaciones correctas: cuanto más oscura la celda, mayor la tasa de acierto.

### A.1. Modelo de recuerdo de voto

La Figura 13 presenta la matriz de confusión del modelo de recuerdo de voto, con una precisión global del 76,9%. El modelo funciona mejor para los partidos mayoritarios: el PP es correctamente clasificado en el 85,6% de los casos y el PSOE en el 83,4%. Vox (79,0%) y los partidos regionales (72,7%) también muestran tasas de acierto elevadas. La abstención se clasifica correctamente en un 62,8% de los casos, con confusiones repartidas principalmente entre PSOE (15,5%) y PP (8,7%), lo cual refleja que el perfil de algunos abstencionistas se solapa con el de votantes poco comprometidos de ambos grandes partidos. Sumar presenta una tasa de acierto del 69,1%, con un 23,5% de sus votantes clasificados como PSOE, lo que es esperable dada la proximidad ideológica entre ambas formaciones. La categoría de «Otros» es la más difícil de predecir (solo un 10,2% de acierto), lo cual tiene sentido: se trata de un grupo heterogéneo por definición, compuesto por votantes de partidos muy diversos que el modelo no puede distinguir entre sí.

## Matriz de confusión: modelo de recuerdo de voto

	PSOE	PP	Abstención	Vox	Sumar	Regionales	Otros
PSOE	83.4	5	4.7	1.5	2.9	2.2	0.3
PP	4.1	85.6	3.3	6.8	0	0.2	0
Abstención	15.5	8.7	62.8	6.5	2.6	3.6	0.3
Vox	2	13.9	4.4	79	0.4	0.4	0
Sumar	23.5	1.5	2.9	1	69.1	2	0
Regionales	14.2	3.4	5.7	1.7	2.3	72.7	0
Otros	22.4	20.4	12.2	10.2	14.3	10.2	10.2

Filas = clase real declarada · Columnas = predicción del modelo · Validación cruzada (5 folds)  
Cada fila suma 100%. La diagonal muestra el porcentaje de aciertos por clase. Precisión global: 76,9%.



**Figura 13.** Matriz de confusión del modelo de recuerdo de voto (validación cruzada, 5 iteraciones). Cada fila representa el recuerdo declarado y cada columna la predicción del modelo. Los porcentajes son relativos a cada fila. Precisión global: 76,9 %.

## A.2. Modelo de intención de voto

La Figura 14 presenta la matriz de confusión equivalente para el modelo de intención de voto, con una precisión global considerablemente mayor: 87,8 %. Esta mejora se explica en gran parte porque el modelo de intención dispone de un predictor adicional muy potente: el recuerdo de voto (ya imputado), que no estaba disponible para el modelo de recuerdo. Los tres partidos mayoritarios (PSOE, Vox, PP) se clasifican con tasas de acierto superiores al 92 %. Los partidos regionales alcanzan un 84,6 % y Sumar un 77,8 %, con el principal error hacia el PSOE (13,2 %), de nuevo por proximidad ideológica. La abstención se clasifica correctamente en un 91,0 % de los casos, lo que indica que los abstencionistas declarados tienen un perfil muy diferenciado del resto. Los partidos más pequeños presentan las tasas más bajas: Podemos (67,2 %), con confusiones hacia Sumar (13,8 %) y PSOE (12,1 %), y SALF (44,9 %), cuyo principal error es hacia Vox (30,6 %). Este último resultado es esperable: SALF es un partido nuevo con pocos casos en la muestra y un perfil de votante que comparte características con el electorado de Vox.

## Matriz de confusión: modelo de intención de voto

Actual	PSOE	PP	Vox	Regionales	Sumar	Abstención	Otros	Podemos	SALF
PSOE	92.7	1	1.5	1.6	1.9	0.9	0.3	0.1	0
PP	2.4	92.3	3.4	0.2	0.2	0.7	0.4	0	0.4
Vox	2.5	3	92.6	0.2	0.2	0.2	0.5	0.2	0.5
Regionales	9.2	1.5	0	84.6	2.6	0.5	1.5	0	0
Sumar	13.2	2.1	1.4	0.7	77.8	1.4	0.7	2.1	0.7
Abstención	2.7	2.7	0.9	1.8	0	91	0.9	0	0
Otros	18.8	10.1	7.2	8.7	1.4	2.9	47.8	1.4	1.4
Podemos	12.1	0	1.7	3.4	13.8	0	1.7	67.2	0
SALF	2	14.3	30.6	2	2	0	4.1	0	44.9

Filas = clase real declarada · Columnas = predicción del modelo · Validación cruzada (5 folds)  
Cada fila suma 100%. La diagonal muestra el porcentaje de aciertos por clase. Precisión global: 87,8%

**Figura 14.** Matriz de confusión del modelo de intención de voto (validación cruzada, 5 iteraciones). Cada fila representa la intención declarada y cada columna la predicción del modelo. Los porcentajes son relativos a cada fila. Precisión global: 87,8 %.

En ambos modelos, los errores de clasificación se concentran entre partidos ideológicamente cercanos (PSOE-Sumar, PP-Vox, Vox-SALF), lo cual es consistente con la estructura del espacio político español y sugiere que el modelo captura correctamente las dimensiones relevantes del comportamiento electoral. Los partidos con menor tasa de acierto son invariablemente los que tienen menos casos en la muestra (Otros, SALF, Podemos), un resultado esperable dado que el modelo dispone de menos información para aprender sus patrones distintivos.

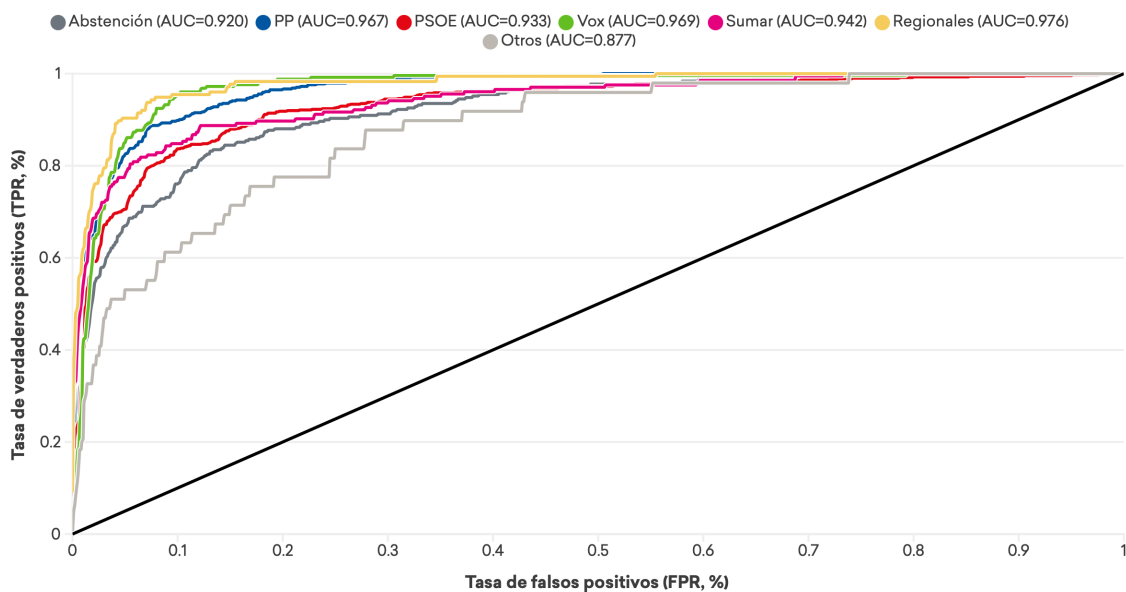
### A.3. Capacidad de discriminación: curvas ROC

Las matrices de confusión miden el acierto del modelo cuando se elige el partido más probable. Pero los modelos XGBoost producen distribuciones de probabilidad completas, y es importante evaluar si esas probabilidades permiten separar bien a los votantes de cada partido del resto del electorado, independientemente del umbral de decisión. Las curvas ROC (*Receiver Operating Characteristic*) ofrecen exactamente esta perspectiva: para cada partido, miden cuánto se solapa la distribución de probabilidades predichas entre quienes realmente votaron a ese partido y quienes no. Un modelo con discriminación perfecta ( $AUC = 1,0$ ) asignaría siempre una probabilidad mayor a los verdaderos votantes que a los no votantes; un modelo aleatorio ( $AUC = 0,5$ ) no distinguiría entre ambos grupos. La Figura 15 muestra las curvas ROC del modelo de recuerdo de voto. Todos los partidos

presentan AUC superiores a 0,87, lo que indica una capacidad de discriminación sólida. Los partidos regionales (AUC = 0,976) son los mejor discriminados, gracias a su perfil geográfico y lingüístico muy distintivo. Vox (0,969) y PP (0,967) se sitúan también muy alto, seguidos por Sumar (0,942) y PSOE (0,933), cuya base más heterogénea hace que sea ligeramente más difícil de distinguir. La abstención (0,920) presenta una discriminación algo menor, coherente con la heterogeneidad de este grupo. La categoría de Otros (0,877) es la que menor AUC obtiene, lo cual es esperable dado que agrupa a votantes de partidos muy diversos sin un perfil común.

### Curvas ROC – recuerdo de voto

Modelo XGBoost · validación cruzada (5 folds)



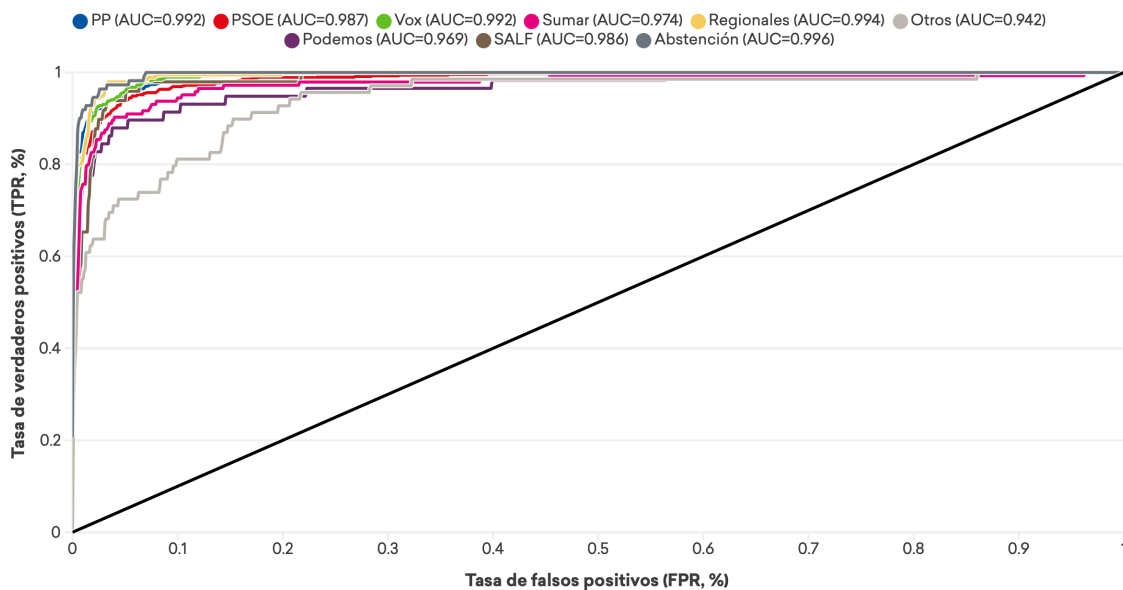
AUC = Area Under Curve.  
Diagonal = clasificador aleatorio (AUC = 0,500).

**Figura 15.** Curvas ROC del modelo de recuerdo de voto (validación cruzada, 5 iteraciones). Cada curva evalúa la capacidad del modelo para separar a los votantes de un partido del resto del electorado. El área bajo la curva (AUC) se indica en la leyenda.

La Figura 16 presenta las curvas equivalentes para el modelo de intención de voto, con AUC notablemente superiores: todos los partidos superan 0,94 y la mayoría están por encima de 0,98. La abstención alcanza un AUC de 0,996, lo que indica que los abstencionistas declarados tienen un perfil casi perfectamente diferenciado. Esta mejora generalizada se explica por la disponibilidad de predictores mucho más informativos: el recuerdo de voto imputado, las baterías de probabilidad de voto (que funcionan como declaraciones suaves de preferencia) y los indicadores de coherencia. Otros (0,942) y Podemos (0,969) siguen siendo las categorías más difíciles de discriminar, pero incluso sus AUC son excelentes.

## Curvas ROC – recuerdo de voto

Modelo XGBoost · validación cruzada (5 folds)



AUC = Area Under Curve.  
Diagonal = clasificador aleatorio (AUC = 0,500).

**Figura 16.** Curvas ROC del modelo de intención de voto (validación cruzada, 5 iteraciones). La discriminación es superior a la del modelo de recuerdo, con todos los AUC por encima de 0,95.

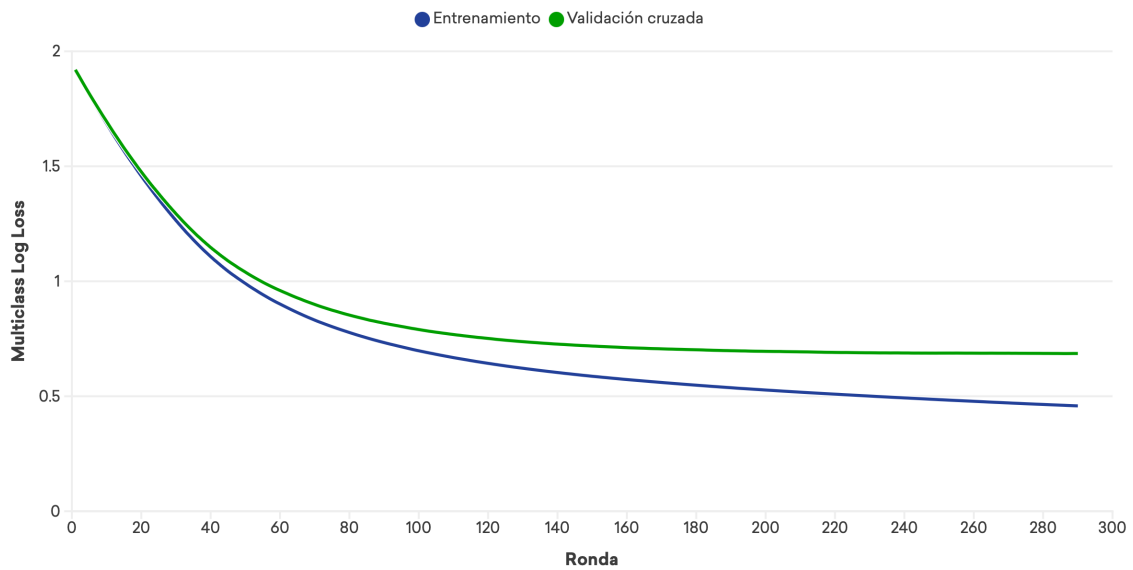
### A.4. Regularización y sobreajuste: curvas de aprendizaje

Un riesgo inherente a los modelos de aprendizaje automático es el sobreajuste: que el modelo memorice los datos de entrenamiento en lugar de capturar patrones generalizables. Las curvas de aprendizaje permiten diagnosticar este problema mostrando cómo evoluciona el error del modelo a medida que se añaden rondas de boosting, tanto sobre los datos de entrenamiento como sobre datos no vistos (validación cruzada).

La Figura 17 muestra la curva de aprendizaje del modelo de recuerdo de voto. La línea de entrenamiento desciende de forma continua a medida que se añaden árboles, mientras que la línea de validación cruzada se estabiliza a partir de la ronda 200 aproximadamente, alcanzando un plateau cerca de 0,69 de log loss multiclase. La brecha entre ambas líneas al final del entrenamiento (290 rondas) es moderada, lo que indica cierto sobreajuste al conjunto de entrenamiento, pero — y esto es lo importante — la curva de validación no sube en ningún momento. Esto confirma que la regularización del modelo (profundidad máxima de 3, submuestreo del 70 %, tasa de aprendizaje de 0,03) es adecuada: los árboles adicionales no perjudican la capacidad de generalización.

## Curva de aprendizaje: modelo de recuerdo de voto

Evolución del error por ronda de boosting



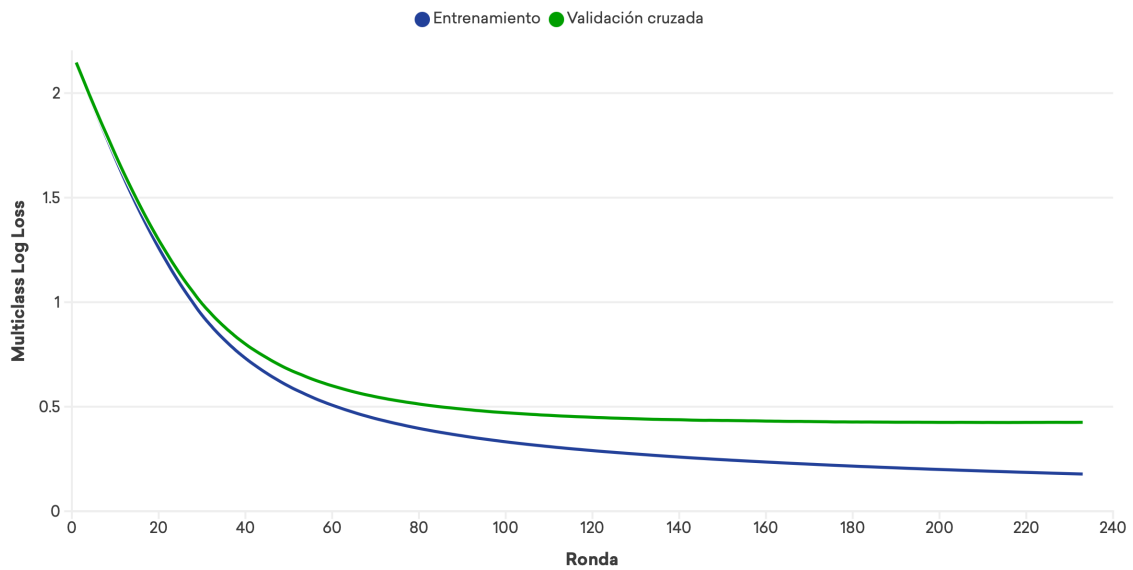
Validación cruzada de 5 iteraciones. La separación entre ambas curvas indica el grado de overfitting.

**Figura 17.** Curva de aprendizaje del modelo de recuerdo de voto. Una línea muestra el error sobre los datos de entrenamiento y la otra sobre validación cruzada. La estabilización de la curva de validación confirma que el modelo no sufre sobreajuste perjudicial.

La Figura 18 presenta la curva equivalente para el modelo de intención de voto, que converge más rápido (plateau a partir de la ronda 150) y alcanza un nivel de error considerablemente menor (log loss  $\approx 0,43$  frente a  $0,69$  en el modelo de recuerdo). Esto es consistente con la mayor calidad de los predictores disponibles. El patrón de regularización es igualmente saludable: la curva de validación se estabiliza sin mostrar indicios de deterioro.

## Curva de aprendizaje: modelo de intención de voto

Evolución del error por ronda de boosting



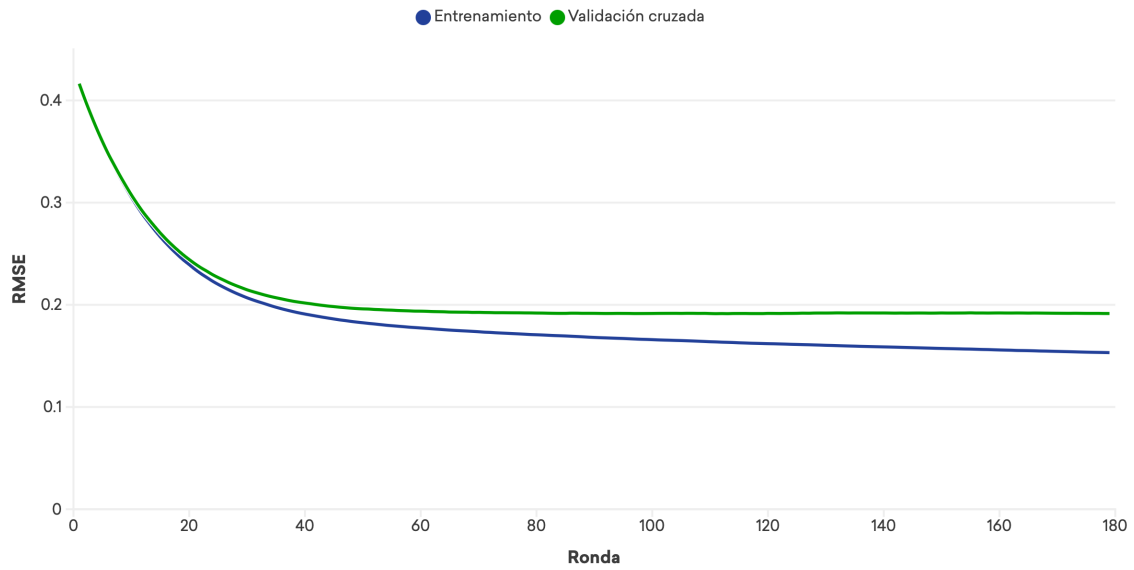
Validación cruzada de 5 iteraciones. La separación entre ambas curvas indica el grado de overfitting.

**Figura 18.** Curva de aprendizaje del modelo de intención de voto. La convergencia es más rápida y el error final menor que en el modelo de recuerdo, reflejando la mayor riqueza de los predictores disponibles.

La Figura 19 muestra la curva de aprendizaje del modelo de participación, que utiliza RMSE como métrica de error (al tratarse de una regresión, no de una clasificación). La convergencia es muy rápida: el modelo se estabiliza alrededor de la ronda 100, con una brecha entre entrenamiento y validación pequeña ( $\approx 0,038$ ). Este modelo (179 rondas, tasa de aprendizaje de 0,05) es coherente con la naturaleza más directa del problema: predecir la propensión a votar a partir de variables observables es una tarea más sencilla que clasificar la preferencia partidista.

## Curva de aprendizaje: modelo de participación

Evolución del error por ronda de boosting



Validación cruzada de 5 iteraciones. RMSE = Root Mean Squared Error.

**Figura 19.** Curva de aprendizaje del modelo de participación (RMSE). La convergencia es rápida y la brecha entre entrenamiento y validación es pequeña, indicando un modelo bien regularizado.

## Bibliografía

---

- Ansolabehere, Stephen y Eitan Hersh (2012). Validation: What Big Data Reveal about Survey Misreporting and the Real Electorate. En: *Political Analysis* 20.4, págs. 437-459.
- Bernstein, Robert, Anita Chadha y Robert Montjoy (2001). Overreporting Voting: Why It Happens and Why It Matters. En: *Public Opinion Quarterly* 65.1, págs. 22-44.
- Blais, André y Carol Galais (2016). Measuring the civic duty to vote: A proposal. En: *Electoral Studies* 41, págs. 60-69.
- Buuren, Stef van (2018). *Flexible Imputation of Missing Data*. 2.<sup>a</sup> ed. CRC Press.
- Cabrera-Álvarez, Pablo y Modesto Escobar (2019). El efecto de la ponderación y la imputación en el sesgo de los estudios electorales en España. En: *Revista Española de Investigaciones Sociológicas* 165, págs. 45-64.
- (2025). Recuerdo de voto y distorsiones analíticas: el caso español tras las elecciones generales de 2023. En: *Revista Española de Investigaciones Sociológicas* 192, págs. 27-46.
- Chen, Tianqi y Carlos Guestrin (2016). XGBoost: A Scalable Tree Boosting System. En: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, págs. 785-794.
- Deming, W. Edwards y Frederick F. Stephan (1940). On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals Are Known. En: *The Annals of Mathematical Statistics* 11.4, págs. 427-444.
- Deng, Yongshi y Thomas Lumley (2023). Multiple Imputation Through XGBoost. En: *Journal of Computational and Graphical Statistics* 33.2, págs. 352-363.
- Deville, Jean-Claude y Carl-Erik Särndal (1992). Calibration Estimators in Survey Sampling. En: *Journal of the American Statistical Association* 87.418, págs. 376-382.
- Durand, Claire, Mélanie Deslauriers e Isabelle Valois (2015). Should Recall of Previous Votes Be Used to Adjust Estimates of Voting Intention? En: *Survey Methods: Insights from the Field*.

- Escobar Mercado, Modesto, Jaime Rivière Gómez y Roberto Cilleros Conde (2014). Los pronósticos electorales con encuestas: elecciones generales en España (1979–2011). Madrid: Centro de Investigaciones Sociológicas.
- Ferrándiz, José Pablo y Francisco Camas García (2019). La cocina electoral en España: La estimación de voto en tiempos de incertidumbre. Madrid: Los Libros de la Catarata.
- Gelman, Andrew (2007). Struggles with survey weighting and regression modeling. En: *Statistical Science* 22.2, págs. 153-164.
- Gelman, Andrew, Sharad Goel, Douglas Rivers y David Rothschild (2016). The Mythical Swing Voter. En: *Quarterly Journal of Political Science* 11.1, págs. 103-130.
- Gill, Jeff (2005). An entropy measure of uncertainty in vote choice. En: *Electoral Studies* 24.3, págs. 371-392.
- Groves, Robert M., Floyd J. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer y Roger Tourangeau (2009). *Survey Methodology*. 2.ª ed. John Wiley & Sons.
- Hastie, Trevor, Robert Tibshirani y Jerome Friedman (2001). *The Elements of Statistical Learning*. Springer.
- Holbrook, Allyson L. y Jon A. Krosnick (2010). Social desirability bias in voter turnout reports: Tests using the item count technique. En: *Public Opinion Quarterly* 74.1, págs. 37-67.
- Instituto Nacional de Estadística (2024). Encuesta de Condiciones de Vida. Consultado en febrero de 2026. URL: [https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica\\_C&cid=1254736176807&menu=ultiDatos&idp=1254735976608](https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176807&menu=ultiDatos&idp=1254735976608).
- (2025). Padrón Continuo. Estadística del Padrón Continuo. Consultado en febrero de 2026. URL: [https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica\\_C&cid=1254736177012&menu=ultiDatos&idp=1254734710990](https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177012&menu=ultiDatos&idp=1254734710990).
- Karp, Jeffrey A. y David Brockington (2005). Social desirability and response validity: A comparative analysis of overreporting voter turnout in five countries. En: *The Journal of Politics* 67.3, págs. 825-840.
- Kennedy, Courtney, Mark Blumenthal, Scott Clement, Joshua D. Clinton, Claire Durand, Charles Franklin, Kyley McGeeney et al. (2018). An Evaluation of the 2016 Election Polls in the United States. En: *Public Opinion Quarterly* 82.1, págs. 1-33.

- Kish, Leslie (1992). Weighting for Unequal  $P_i$ . En: Journal of Official Statistics 8.2, págs. 183-200.
- Little, Roderick J. A. y Donald B. Rubin (2019). Statistical Analysis with Missing Data. 3.<sup>a</sup> ed. John Wiley & Sons.
- Llaneras, Kiko (oct. de 2025). ¿Hay que ignorar las estimaciones de voto del CIS? En: El País. URL: <https://elpais.com/espana/2025-10-14/hay-que-ignorar-las-estimaciones-de-voto-del-cis.html>.
- Lumley, Thomas (2010). Complex Surveys: A Guide to Analysis Using R. John Wiley & Sons.
- Maldita.es (2019). ¿Qué es la famosa cocina del CIS? Consultado en marzo de 2026. URL: <https://maldita.es/malditateexplica/20190412/que-es-la-famosa-cocina-del-cis/>.
- Ministerio del Interior (2023). Resultados electorales. Elecciones generales 23 de julio de 2023. Consultado en febrero de 2026. URL: <https://resultados.generales23j.es/>.
- Morin-Chassé, Alexandre, Damien Bol, Laura B. Stephenson y Simon Labbé St-Vincent (2017). How to Survey About Electoral Turnout? The Efficacy of the Face-Saving Response Items in 19 Different Contexts. En: Political Science Research and Methods 5.3, págs. 575-584.
- Pennay, Darren, Sebastian Misson, Dina Neiger y Paul J. Lavrakas (2023). How Weighting by Past Vote Can Improve Estimates of Voting Intentions. En: Survey Practice 16.1.
- Pew Research Center (2016). Can Likely Voter Models Be Improved? Consultado en marzo de 2026.
- (2018). How Different Weighting Methods Work. Consultado en marzo de 2026.
- Rivero, Gonzalo y Jordi Muñoz (2022). La metodología de las estimaciones electorales del CEO. Papers de treball del Centre d'Estudis d'Opinió, versió 1.3.
- Rubin, Donald B. (1987). Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons.
- Selb, Peter y Simon Munzert (2013). Voter overrepresentation, vote misreporting, and turnout bias in postelection surveys. En: Electoral Studies 32.1, págs. 186-196.

Shirani-Mehr, Houshmand, David Rothschild, Sharad Goel y Andrew Gelman (2018). Disentangling Bias and Variance in Election Polls. En: Journal of the American Statistical Association 113.522, págs. 607-614.

Urquizu, Ignacio (2005). El voto oculto en España. En: Revista Española de Ciencia Política 13, págs. 119-156.