# Variables sociales y voto por el PSOE en 2023

Las variables determinantes del voto de cada individuo pueden agruparse en tres niveles básicos: idiosincráticas, familiares y sociales. El peso de cada una de ellas en la decisión electoral puede variar, pero su influencia general puede ser determinada para diferentes grupos. El objetivo de este estudio es evaluar la influencia de un conjunto de variables en la determinación de la probabilidad de votar por el Partido Obrero Socialista Español (PSOE) en 2023.

# Metodología

## Problema de investigación

El problema de investigación puede ser definido de la manera siguiente: bajo la premisa de que cada individuo, en tanto que miembro de un grupo social base (amigos, condiscípulos, colegas, etc.) tiende a unirse con personas de características semejantes, ¿cuál es la influencia de estos rasgos en su decisión de voto por el PSOE? En particular, éste estudio se centrará evaluar la influencia de sexo, edad, nivel educativo, ingreso y religión.

## Hipótesis y variables

### Hipótesis nula

La probabilidad de voto por el PSOE no está significativamente asociada con el sexo, la edad, el nivel educativo, el ingreso o la religión.

### Hipótesis alterna

La probabilidad de voto por el PSOE está significativamente asociada con el sexo, la edad, el nivel educativo, el ingreso y la religión. La forma específica de la asociación es:

* Intención de voto por el PSOE. Variable dependiente, dicotómica.
* Sexo: Hay una diferencia significative entre hombres y mujeres, de forma tal que los hombres votan más por el PSOE. Variable independiente, dicotómica.
* Edad: Los votantes más jóvenes son más propensos a votar por el PSOE que los de mayor edad. Variable independiente, continua.
* Nivel educativo: Las personas con un mayor nivel educativo votan más por el PSOE. Variable independiente, continua.
* Ingreso: Un menor nivel de ingresos está asociado con una mayor probabilidad de voto por el PSOE. Variable independiente, continua.
* Religión: Los católicos tiene una menor probabilidad de votar por el PSOE. Variable independiente, dicotómica.

## Fuente de los datos

Para la evaluación de la hipótesis usamos los datos del *Barómetro 3413* de julio de 2023 del Centro de Investigaciones Sociológicas (CIS) (2023b). El estudio original contenía 33 preguntas que fueron respondidas por 4.164 entrevistados. Para esta investigación se tomaron en cuenta las preguntas relevantes sobre intención de voto en las elecciones generales de 2023 (P7), sexo (P0a), edad (P0b), nivel educativo (P18a), religión (P19) e ingresos del hogar (P23). Las variables fueron codificadas y se seleccionaron únicamente los casos completos (n = 3.951). El análisis se hizo con el programa R (2021) y las siguientes librerías: Car, Tidyverse y Writexl. Para reproducibilidad, las funciones y operaciones usadas en las diferentes etapas pueden descargarse del correspondiente repositorio de GitHub (Ortiz, 2025).

# Estadística descriptiva

Descripción de las variables

Las variables independientes usadas pueden separarse en dos categorías: cualitativas y dicotómicas. Entre las variables cuantitativas, la edad promedio es un poco más de 50 años 3 meses con un desvío estándar de 16 años y 6 meses. El nivel educativo promedio es 8,09 (estudios técnicos superiores a FP e inferiores a grado) (CIS, 2023a, p. 5). El ingreso por hogar promedio está dentro del nivel 3 (de 1.801 a 2.700 €) (CIS, 2023a, p. 6) (ver Tabla 1). Ambas variables dicotómicas muestran una distribución homogénea cercana al 50 por ciento para cada uno de los grupos. Así, las mujeres son un 48,62 por ciento de los encuestados y los católicos el 48,32 por ciento (ver Tabla 2).

Tabla 1  
*Estadísticos descriptivos de las variables cuantitativas*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | M | SD | Min | Max |
| EDAD | 50,27 | 16,53 | 18 | 93 |
| NIVELESTENTREV | 8,09 | 3,39 | 1 | 15 |
| INGRESHOG | 3,46 | 1,46 | 1 | 6 |
| n = 3.951 |  |  |  |  |
|  |  |  |  |  |

Tabla 2  
*Estadísticos descriptivos de las variables dicotómicas*

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Valores | n | Porcentaje |
| SEXO | Mujer | 1921 | 48,62 |
|  | Hombre | 2030 | 51,38 |
| RELIGION | No católico | 1909 | 48,32 |
|  | Católico | 2042 | 51,68 |
| n = 3.951 |  |  |  |
|  |  |  |  |

## Análisis de multicolinealidad

Previo a la operación de regresión logística, es necesario conocer si las variables independientes están altamente correlacionadas entre sí. De ser éste el caso, la validez de los resultados pudiera verse afectada (Schreiber-Gregory, 2017, p. 1). La medida usada en este caso es el factor de inflación de la varianza (VIF), la cual mide el nivel en que la varianza del coeficiente de regresión de un predictor está inflada por su correlación con los otros predictores. Este estadístico oscila entre 1 e infinito y cualquier valor por debajo de 4 se considera aceptable (Comprehensive R Archive Network [CRAN]). Como se puede observar en la Tabla 3, ninguna de las variables dependientes alcanza un nivel de riesgo para nuestra investigación.

Tabla 3  
*Factor de inflación de la varianza*

|  |  |
| --- | --- |
| Variable | VIF |
| SEXO | 1,02 |
| EDAD | 1,11 |
| RELIGION | 1,08 |
| NIVELESTENTREV | 1,29 |
| INGRESHOG | 1,24 |
|  |  |

# Resultados

Luego de estos análisis preliminares llegamos a la operación principal, la regresión logística. La eliminación hacia atrás (*backward stepwise regression*) fue usada para la selección de las variables en el modelo final. Éstas deben alcanzar un nivel de significación menor a 0,01 (*p < 0,01*) para poder ser seleccionadas. En el modelo inicial (Tabla 4) se incluyeron todas las variables independientes y sus interacciones. Para los subsiguientes modelos se incluyeron las variables independientes significativas, las variables independientes presentes en interacciones significativas y las interacciones significativas. Por ejemplo, el modelo completo muestra que el ingreso por hogar (INGRESHOG) no tiene un valor *p* significativo (*p > ,01*), sin embargo, su interacción con sexo sí lo es. Por ello, ingreso por hogar fue incluido en los modelos subsiguientes.

Tabla 4  
*Modelo Completo*

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | co | Error S. | Odds Ratio |
| (Intercept) | -1,84\*\* | 0,57 | 0,16 |
| SEXO | -1,93\*\*\* | 0,40 | 0,15 |
| EDAD | 0,02\* | 0,01 | 1,02 |
| RELIGION | 0,32 | 0,40 | 1,40 |
| NIVELESTENTREV | 0,07 | 0,05 | 1,10 |
| INGRESHOG | 0,07 | 0,14 | 1,07 |
| SEXO:EDAD | 0,02\*\* | 0,00 | 1,02 |
| SEXO:RELIGION | -0,07 | 0,17 | 0,93 |
| SEXO:NIVELESTENTREV | 0,08\*\* | 0,03 | 1,08 |
| SEXO:INGRESHOG | -0,01 | 0,10 | 0,99 |
| EDAD:RELIGION | -0,01 | 0,00 | 0,99 |
| EDAD:NIVELESTENTREV | -0,00\*\* | 0,00 | 1,00 |
| EDAD:INGRESHOG | 0,00 | 0,00 | 1,00 |
| RELIGION:NIVELESTENTREV | -0,00 | 0,03 | 1,00 |
| RELIGION:INGRESHOG | -0,16\* | 0,06 | 0,85 |
| NIVELESTENTREV:INGRESHOG | -0,00 | 0,01 | 1,00 |
| *\*p > ,01; \*\*p > ,001; \*\*\*p > 0* |  |  |  |
|  |  |  |  |

Tabla   
*Modelo final*

|  |  |  |  |
| --- | --- | --- | --- |
| Variables | co | Error S. | Odds Ratio |
| (Intercept) | -0,76\*\* | 0,25 | 0,46 |
| SEXO | -1,83\*\*\* | 0,36 | 0,16 |
| EDAD | 0,00 | 0,00 | 1,00 |
| NIVELESTENTREV | -0,04\*\* | 0,02 | 0,96 |
| SEXO:EDAD | 0,02\*\*\* | 0,01 | 1,02 |
| SEXO:NIVELESTENTREV | 0,06\*\* | 0,02 | 1,06 |
| *\*p > ,01; \*\*p > ,001; \*\*\*p > 0* |  |  |  |
|  |  |  |  |

# Conclusión

# Referencias

Centro de Investigaciones Sociológicas. (2023a). *3413/0 Barómetro de julio 2023* [Códigos de encuesta]. Centro de Investigaciones Sociológicas. [**https://www.cis.es/es/detalle-ficha-estudio?idEstudio=14729**](https://www.cis.es/es/detalle-ficha-estudio?idEstudio=14729).

Centro de Investigaciones Sociológicas. (2023b). *3413/0 Barómetro de julio 2023* [Conjunto de datos]. Centro de Investigaciones Sociológicas. [**https://www.cis.es/es/detalle-ficha-estudio?idEstudio=14729**](https://www.cis.es/es/detalle-ficha-estudio?idEstudio=14729).

Comprehensive R Archive Network. (n.d.). *Collinearity Diagnostics, Model Fit & Variable Contribution*. Cran.r-project.org. **https://cran.r-project.org/web/packages/olsrr/vignettes/regression\_diagnostics.html**

Ortiz, Marco. (2025). *ensayo\_regresion\_logistica.R*. GitHub. **https://github.com/ortizpalanques1/CIS/blob/main/ensayo\_regresion\_logistica.R**.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. [**https://www.R-project.org/**](https://www.R-project.org/).

Schreiber-Gregory, Deanna. (2017). Multicollinearity: What Is It, Why Should We Care, and How Can It Be Controlled? *Proceedings SAS Global Forum.* **https://support.sas.com/resources/papers/proceedings17/1404-2017**.pdf.