



Universidad de Valladolid

FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

Trabajo Fin de Grado

Grado en Marketing e Investigación de Mercados

Un Modelo Logit para el Análisis de los Determinantes del Desempleo

Presentado por:

Daniel Rodríguez Fernández

Valladolid, 14 de julio de 2025

RESUMEN

En este trabajo vamos a tratar de concretar como afectan características personales de los individuos como su edad, sexo, estado civil, nacionalidad, la comunidad autónoma en la que residen o su nivel de formación a su situación en el mercado laboral, es decir, si están ocupados, o, por el contrario, se encuentran en situación de desempleo. Para ello usaremos la encuesta de la EPA realizada por el Instituto Nacional de Estadística correspondiente al primer trimestre del año 2.025.

El análisis de estos datos consiste en un primer análisis descriptivo de las diferentes variables, para posteriormente construir un modelo de regresión logística o modelo logit con el que poder estimar la probabilidad de estar desempleado en función de las características de un individuo.

Gracias a estos análisis hemos podido concluir que el principal determinante de la situación laboral es el nivel de formación, aunque otros determinantes como la edad, el estado civil y la nacionalidad también marcan importantes diferencias.

ABSTRACT

In this study, we aim to determine how personal characteristics of individuals such as their age, sex, marital status, nationality, the autonomous community in which they reside, and their level of education affect their situation in the labor market, that is, whether they are employed or, on the contrary, unemployed. To this end, we use data from the Labor Force Survey (EPA) conducted by the Spanish National Statistics Institute (INE) for the first quarter of 2025.

The analysis of these data begins with a descriptive exploration of the different variables, followed by the construction of a logistic regression model (logit model) that allows us to estimate the probability of being unemployed based on an individual's characteristics.

Thanks to these analyses, we have been able to conclude that the main determinant of employment status is the level of education, although other factors such as age, marital status, and nationality also produce significant differences.

ÍNDICE

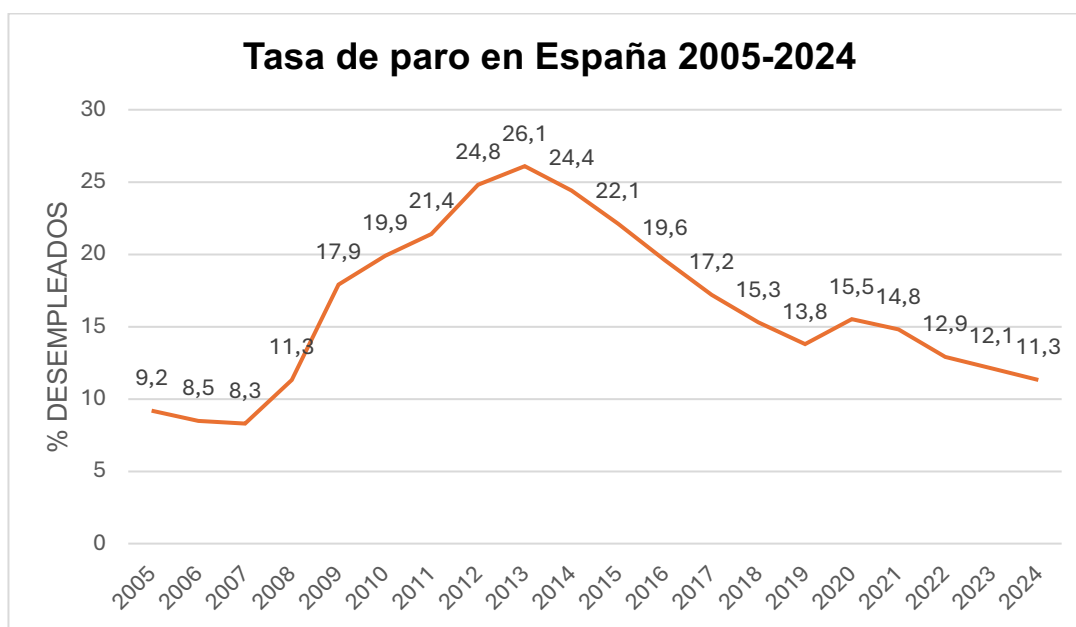
1. INTRODUCCIÓN	7
2. TRATAMIENTO DE DATOS.....	8
2.1. Metodología	8
2.2. Análisis descriptivo	9
2.2.1. Análisis de la variable EDAD	11
2.2.2. Análisis de la variable CC.AA.	12
2.2.3. Análisis de la variable SEXO	14
2.2.4. Análisis de la variable ESTCIVIL	14
2.2.5. Análisis de la variable FORMACION	16
3. MODELO LOGIT.....	18
3.1. Marco teórico	18
3.2. Análisis del modelo logit	21
3.2.1. Efecto marginal de cada variable	23
4. CREACIÓN Y ANÁLISIS DE LOS ESCENARIOS	26
4.1. Curvas de respuesta	31
5. CONCLUSIONES.....	35
6. BIBLIOGRAFÍA	37

Tabla 1. Frecuencia y porcentaje general de parados y ocupados.	11
Tabla 2. Estimación del modelo logarítmico	22
Tabla 3. Clasificación del modelo logarítmico	23
Tabla 4. Efecto marginal de la variable EDAD	24
Tabla 5. Efecto marginal de la variable FORMACIÓN	25
Tabla 6. Efecto marginal de las variables SEXO, ESTCIVIL y NACIONALIDAD.....	25
Tabla 7. Escenarios	26
Ilustración 1. Evolución de la tasa de paro en España	7
Ilustración 2. Porcentaje (%) de desempleo en función de la variable EDAD	12
Ilustración 3. Porcentaje (%) de desempleo en función de la variable CC.AA.	13
Ilustración 4. Porcentaje (%) de desempleo en función de la variable SEXO.	14
Ilustración 5. Porcentaje (%) de desempleo en función de la variable ESTCIVIL.	15
Ilustración 6. Porcentaje (%) de desempleo en función de la variable ESTCIVIL según la variable SEXO.	16
Ilustración 7. Porcentaje (%) de desempleo en función de la variable FORMACION. ...	17
Ilustración 8. Porcentaje (%) de desempleo en función de la variable NACIONALIDAD.	18
Ilustración 9. Gráfico curva EDAD-SEXO	32
Ilustración 10. Gráfico curva EDAD-ESTADO CIVIL	33
Ilustración 11. Gráfico curva EDAD-NACIONALIDAD	33
Ilustración 12. Gráfico curva EDAD-FORMACIÓN	34

1. INTRODUCCIÓN

El desempleo ha sido históricamente uno de los principales problemas económicos y sociales en España. Desde los años ochenta, la tasa de paro española rara vez ha descendido del 8 %, manteniéndose por encima de la media europea y posicionando a España de forma recurrente entre los países con mayores tasas de desempleo de la UE y la OCDE. Este fenómeno alcanzó su punto crítico durante la crisis financiera, cuando en el primer trimestre de 2013 el desempleo superó el 26 %.

Ilustración 1. Evolución de la tasa de paro en España



Fuente: Elaboración propia a partir de datos del INE

Aunque en la última década la evolución ha sido favorable, con una clara tendencia descendente, el mercado laboral español todavía acusa debilidades estructurales. En 2024, la tasa de desempleo promedio se situó en torno al 11,3 %, y si bien España lideró el crecimiento económico entre las economías avanzadas con un aumento del PIB del 3,2 %, persistieron retos como la elevada temporalidad, la precariedad y el elevado desempleo juvenil. Este último sigue siendo particularmente preocupante, con una tasa para menores de 25 años cercana al 26 %, más del doble que la media europea.

En este contexto, el presente trabajo se centra en analizar la situación del desempleo en España durante el primer trimestre de 2025, periodo en el que, tras varios trimestres de

mejora, se registró un repunte del desempleo hasta el 11,36 %, el mayor incremento inter-trimestral desde 2013, afectando a cerca de 2,79 millones de personas. No obstante, en términos interanuales, la tendencia sigue siendo positiva, con una caída de casi un punto respecto al año anterior.

El objetivo de este estudio es profundizar en los factores personales que determinan la probabilidad de estar desempleado, evaluando cómo influyen variables como el sexo, la edad, el estado civil, la nacionalidad y el nivel educativo en la situación laboral. Para ello, se parte de un análisis descriptivo inicial que permite visualizar las desigualdades existentes entre grupos poblacionales, para posteriormente estimar un modelo logit que cuantifica la incidencia de cada uno de estos factores sobre la probabilidad de desempleo.

Este enfoque facilita no solo corroborar los patrones detectados a nivel descriptivo, sino también simular escenarios con combinaciones específicas de características individuales, identificando así los perfiles más vulnerables al desempleo. Con ello, se pretende aportar evidencia empírica que pueda servir de apoyo para diseñar políticas públicas orientadas a reducir el paro estructural, incidir especialmente en el desempleo juvenil y mejorar la inserción laboral de los colectivos con menor nivel de cualificación.

2. TRATAMIENTO DE DATOS

2.1. Metodología

Para realizar esta investigación se escogió una muestra correspondiente al primer trimestre del año 2025 de la EPA realizada por el INE.

El motivo de elegir este periodo es estudiar los datos más recientes disponibles, para tratar de analizar los determinantes del desempleo en la coyuntura económica actual.

Para realizar estos análisis se ha empleado el software estadístico SPSS, tanto para el análisis descriptivo de las variables escogidas como determinantes del desempleo de interés para el estudio como para la construcción del modelo logarítmico, y el programa Excel para el filtrado de la base de datos y la elaboración de todos los gráficos de la investigación, así como para el cálculo de las probabilidades obtenidas en el apartado de los escenarios.

2.2. Análisis descriptivo

Para la construcción del modelo logístico, se eligieron las variables que pudieran representar los determinantes más importantes a la hora de determinar las causas del desempleo.

Se seleccionaron las siguientes variables:

SEXO1: Hace referencia al género del individuo encuestado. Se codifica originalmente como: Hombre = 1, Mujer = 6. Se recodifica como: Hombre = 1; Mujer = 0. La variable pasa a denominarse **SEXO**.

CCAA: Indica la comunidad autónoma en la que reside el individuo encuestado. La variable numera del 01 al 17 los territorios por orden alfabético, con la excepción de Ceuta y Melilla, que se codifican como: 51 = Ceuta y 52 = Melilla. La variable se nombra como **CCAA**.

EDAD1: Representa la edad del individuo en rangos de cinco años. Representando el intervalo de edad con el número más pequeño del rango. Por ejemplo, para el intervalo de individuos de entre 25 a 29 años, en la base de datos se representa con un 25. La variable se renombra como **EDAD**.

ECIV1: Recoge el estado civil de los individuos encuestados. Originalmente la variable se codifica como: 1: Soltero; 2: Casado; 3: Viudo; 4: Separado o Divorciado. Para realizar el análisis descriptivo inicial se mantuvieron estos grupos, pero, posteriormente con el objetivo de facilitar la construcción del modelo logarítmico la variable se agrupó y recodificó de la siguiente forma:

- 0: Casado
- 1: Soltero. Que contiene los grupos 1: soltero, 3: viudo y 4: separado o divorciado.

La variable pasa a denominarse **EST.CIVIL**.

NFORMA: Recoge el nivel de formación más alto alcanzado por el individuo. Debido a la gran variedad de opciones se recodifica la variable para mayor sencillez tanto en el análisis como en los resultados, por lo que se recodifica la variable de la siguiente forma: 1. Primaria; 2: Secundaria; 3: Bachillerato; 4: Estudios Superiores. La variable pasa a denominarse como **FORMACION**.

NAC1: Indica la nacionalidad de los individuos de la muestra. Se codifica originalmente como: 1: española; 2: española y doble nacionalidad; 3: extranjera. Se agrupó y recodificó de la siguiente forma:

- 1: española. Que contiene los grupos 1: española y 2: española y doble nacionalidad.
- 0: extranjera

La variable pasa a denominarse **NACIONALIDAD**.

AOI: Se trata de la variable principal de este estudio. Según la propia base de datos de la EPA que se utilizó la define como: la clasificación de los entrevistados por relación con la actividad económica según criterios OIT. La propia base de datos utiliza la siguiente clasificación:

- 03: Ocupados subempleados por insuficiencia de horas.
- 04: Resto de ocupados.
- 05: Parados que buscan primer empleo.
- 06: Parados que han trabajado antes.
- 07: Inactivos 1 (desanimados).
- 08: Inactivos 2 (junto con los desanimados forman los activos potenciales)
- 09: Inactivos 3 (resto de inactivos)

Como para este estudio lo que se buscó fue analizar los determinantes del desempleo, en primer lugar, se eliminaron de la base de datos todos aquellos casos de encuestados clasificados como inactivos, que, aunque están desempleados se encuentran en una búsqueda activa de empleo.

Para una mayor sencillez tanto en el primer análisis descriptivo como para la posterior construcción del modelo logístico, los cuatro grupos de encuestados resultantes (03, 04, 05, 06) se agrupan en recodifican en dos nuevos grupos:

- 0: Ocupados. Que contiene los grupos 03 y 04, ocupados subempleados y resto de ocupados
- 1: Parados. Que contiene los grupos 05 y 06, parados que buscan primer empleo y parados que han trabajado antes.

A continuación, para comenzar el estudio, se realizó un análisis descriptivo de la muestra, con lo que se comprobó el nivel de desempleo para cada uno de los determinantes (variables) escogidas, como, por ejemplo, las diferencias en función de la edad, o del estado civil. También se observó cómo afectan diferentes variables juntas, por ejemplo, el desempleo entre hombres casados y solteros, y mujeres casadas y solteras.

Antes de realizar el análisis con cada variable, se obtuvo el porcentaje de desempleo general de la muestra, para posteriormente poder hacer comparaciones con cada una de las variables.

Tabla 1. Frecuencia y porcentaje general de parados y ocupados.

AOI		Frecuencia	Porcentaje
Válido	OCUPADOS	51053	89,2
	PARADOS	6168	10,8
	Total	57221	100,0

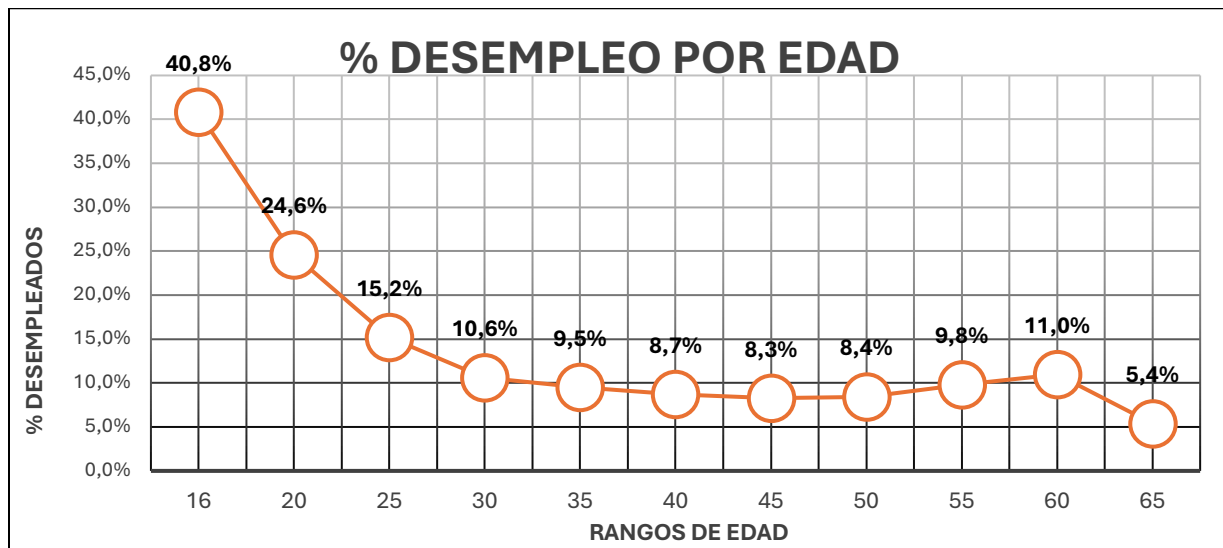
Fuente: Elaboración propia con SPSS desde datos EPA

En el análisis realizado se obtuvo, para el global de la muestra, una tasa general del desempleo del 10,8%.

2.2.1. Análisis de la variable EDAD

Se comenzó con la variable EDAD, que agrupa la edad en rangos de cinco años, representándolos con la edad menor. En el gráfico se observa una tendencia clara en el nivel de desempleo según la edad.

Ilustración 2. Porcentaje (%) de desempleo en función de la variable EDAD



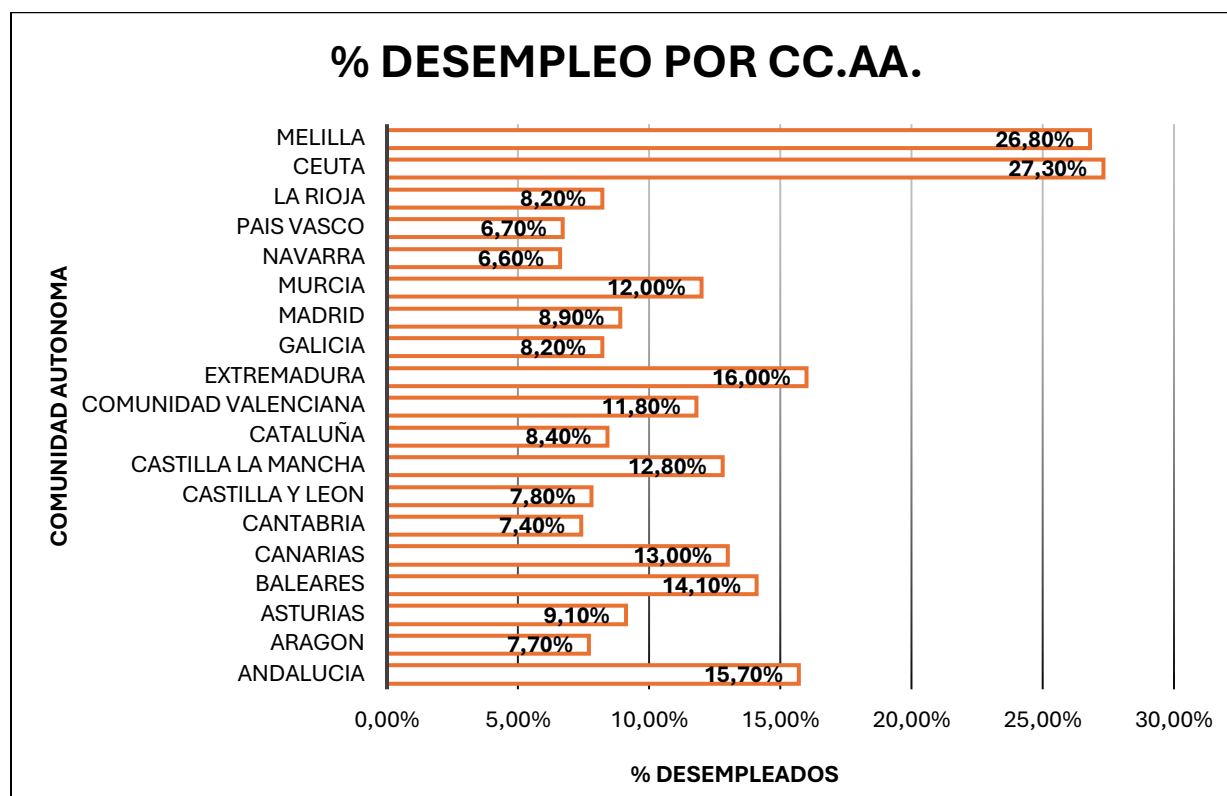
Fuente: Elaboración propia con Excel desde datos EPA

Gracias a este análisis se obtuvo que el nivel de desempleo está claramente condicionado por la edad del individuo. Obteniendo una tendencia clara, en la que entre los individuos de mayor edad existe un menor número de desempleados. Es especialmente llamativo el caso de los jóvenes, para los individuos de entre 20 y 25 años, se obtuvo una tasa de paro de cerca del 25%. Entre los 25 y los 55 años se estabiliza, situándose la tasa por debajo de la media general de la muestra. Y entre los 55 y los 60 años el porcentaje de desempleados vuelve a aumentar de nuevo hasta situarse por encima del 10%. Para que posteriormente al llegar a los 65, la edad de jubilación, el porcentaje de desempleados se desplome hasta un 5,4%.

2.2.2. Análisis de la variable CC.AA.

La siguiente variable analizada fue CC. AA, con el objetivo de tratar de identificar si existen diferencias significativas en función de la zona geográfica en la que residen los individuos de la muestra.

Ilustración 3. Porcentaje (%) de desempleo en función de la variable CC.AA.



Fuente: Elaboración propia con Excel desde datos EPA

Se observó que, si existen diferencias claras entre las diferentes comunidades pudiendo agruparlas, según su porcentaje de desempleados obtenido, de la siguiente forma:

Grupo 1: Compuesto por las CC. AA con una tasa de desempleo menor al 10%.

- Principado de Asturias, Comunidad de Madrid, Comunidad Foral de Navarra, País Vasco, Cantabria, Galicia, La Rioja, Cataluña, Aragón, Castilla y León.

Grupo 2: Correspondiente con las CC. AA con una tasa de desempleo de entre el 10 y el 20%.

- Andalucía, Extremadura, Comunidad Valenciana, Murcia, Canarias, Castilla la Mancha, Islas Baleares.

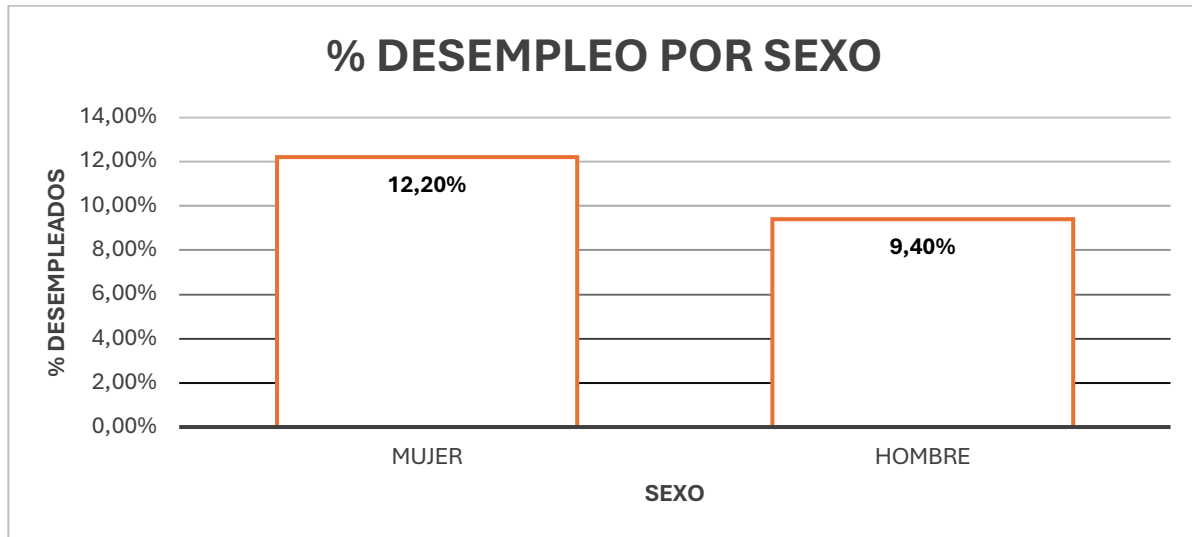
Grupo 3: Compuesto por las CC. AA con una tasa de desempleo superior al 20%.

- Ceuta, Melilla

2.2.3. Análisis de la variable SEXO

Se analizó la variable SEXO para comprobar si existen diferencias significativas en la tasa de paro entre hombres y mujeres.

Ilustración 4. Porcentaje (%) de desempleo en función de la variable SEXO.



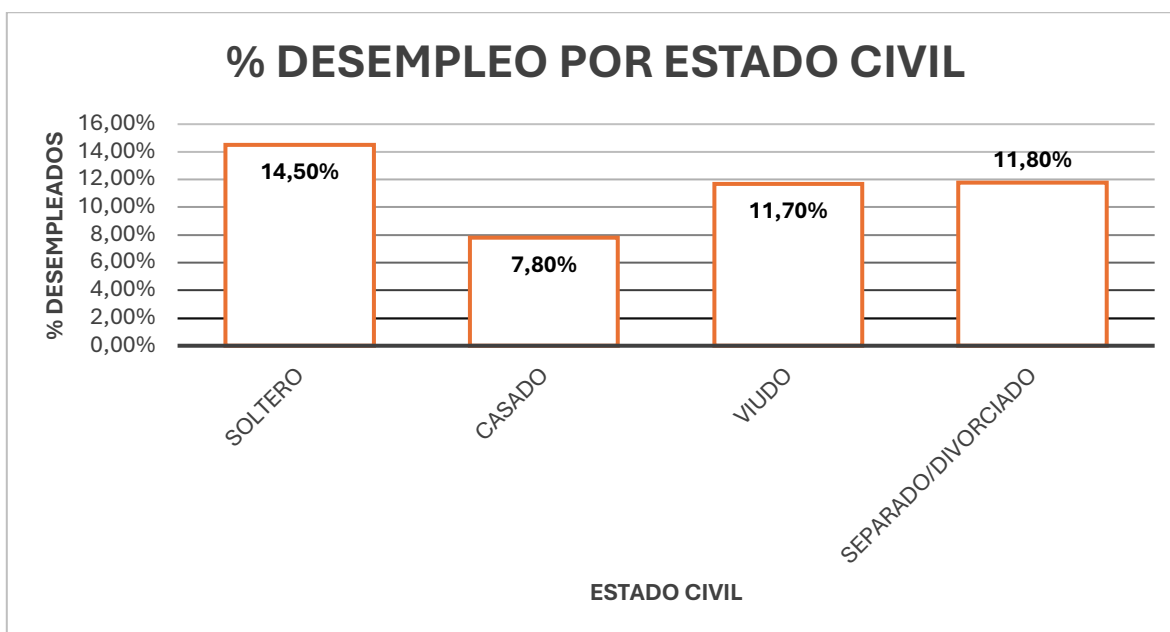
Fuente: Elaboración propia con Excel desde datos EPA

Con los resultados obtenidos se observó una diferencia de cerca de 3 puntos porcentuales en las tasas de paro obtenidas, resultando un 9,40% para los hombres, y un 12,2% para las mujeres.

2.2.4. Análisis de la variable ESTCIVIL

Se analizó la variable ESTCIVIL para comprobar si existen diferencias en la tasa de paro en función del estado civil de las personas.

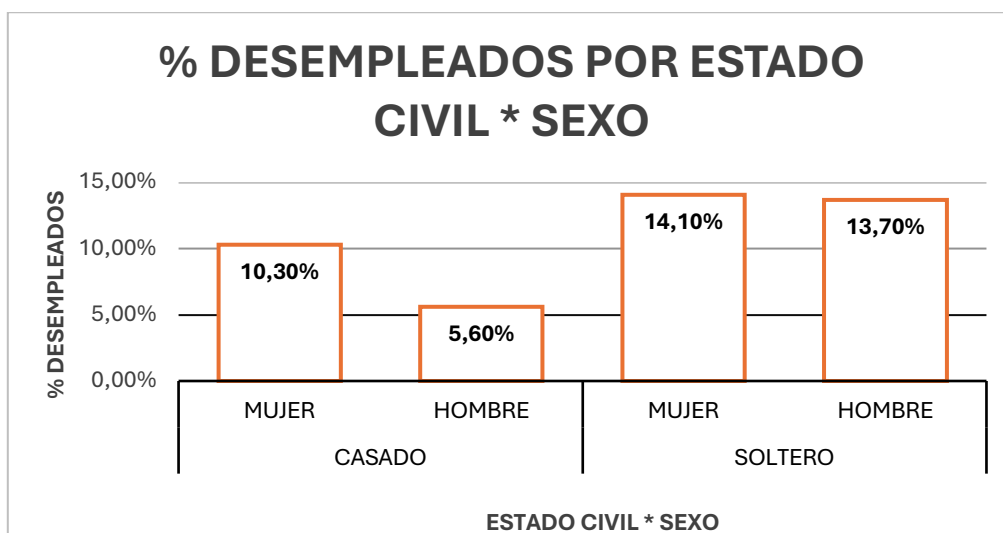
Ilustración 5. Porcentaje (%) de desempleo en función de la variable ESTCIVIL.



Fuente: Elaboración propia con Excel desde datos EPA

En los resultados que se obtuvieron se observa una diferencia bastante significativa, los individuos solteros cuentan con una tasa de desempleo del 14,5%, muy superior a la de los individuos casados que únicamente el 7,8% se encuentra en una situación de desempleo. Por otro lado, los individuos viudos o separados/divorciados cuentan con una tasa de paro muy similar, 11,7% y 11,8% respectivamente. Esto podría deberse a diferentes factores, aunque si observamos las variables analizadas anteriormente podríamos relacionarlo con el desempleo según la EDAD, ya que como es lógico pensar, las personas más jóvenes, que son las de una mayor tasa de paro según su edad, que sería el caso de los solteros. Mientras que las personas de una mediana edad, y, por tanto, con una menor tasa de paro según su edad, por lo general son personas ya casadas. Para el caso de los individuos viudos o separados/divorciados sería interesante buscar el motivo del incremento de la tasa de paro, que puede estar relacionado con la edad, o la situación familiar.

Ilustración 6. Porcentaje (%) de desempleo en función de la variable ESTCIVIL según la variable SEXO.



Fuente: Elaboración propia con Excel desde datos EPA

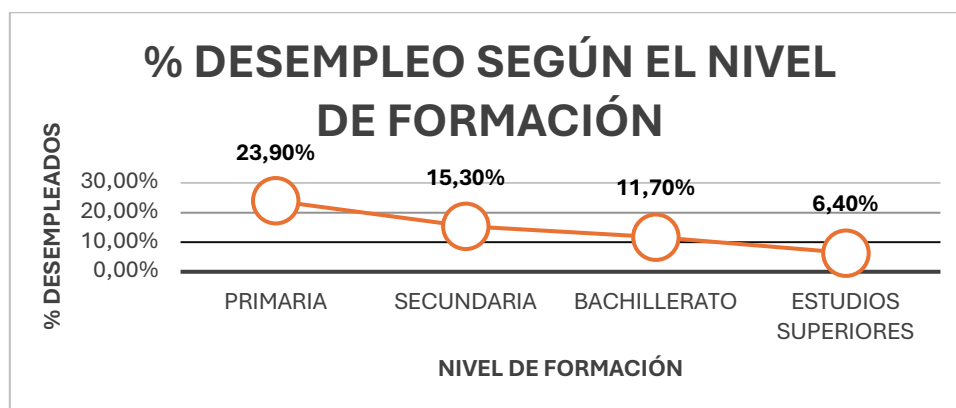
También se analizó como afectaba las variables correspondientes al estado civil y al sexo de forma conjunta. Se observó que para los individuos solteros la tasa de desempleo es ligeramente superior en el caso de las mujeres, aunque nada significativo, únicamente 0,4 puntos porcentuales; como se analizó en la variable del estado civil, seguramente condicionado también por la edad.

En el caso de los individuos casados, sí que observaron diferencias significativas, para los hombres el porcentaje de desempleo se sitúa en un 5,6% marcando una diferencia de casi 5 puntos porcentuales con las mujeres que se sitúan con apenas diferencia con la tasa general, con un 10,3%. Esto puede deberse a que, a las mujeres casadas, aun se las relaciona y se las otorga el cuidado de los hijos y el hogar familiar, lo que en muchos casos no las permite incorporarse al mercado laboral por falta de tiempo, incrementando así la diferencia en el porcentaje de desempleo con los hombres.

2.2.5. Análisis de la variable FORMACION

Se estudió la variable FORMACION, para analizar las diferencias en la tasa de paro de los individuos en función del diferente nivel de estudios alcanzado por los individuos.

Ilustración 7. Porcentaje (%) de desempleo en función de la variable FORMACION.



Fuente: Elaboración propia con Excel desde datos EPA

Tras el análisis se observó, como era de esperar, que los individuos con un menor nivel de estudios finalizado son los que tienen una mayor probabilidad de no estar ocupados, siendo del 23,9% para los individuos que como máximo finalizaron la educación primaria, y del 15,3% para los que como máximo terminaron la educación secundaria.

El siguiente escalón (bachillerato) lo forman los individuos que cuentan con el título de bachillerato o que han finalizado un grado medio o estudios equivalentes, para este grupo de individuos, el porcentaje de desempleo se reduce hasta el 11,7%, apenas un punto porcentual por encima de la media, lo que podríamos comenzar a considerar como una tasa de paro normal.

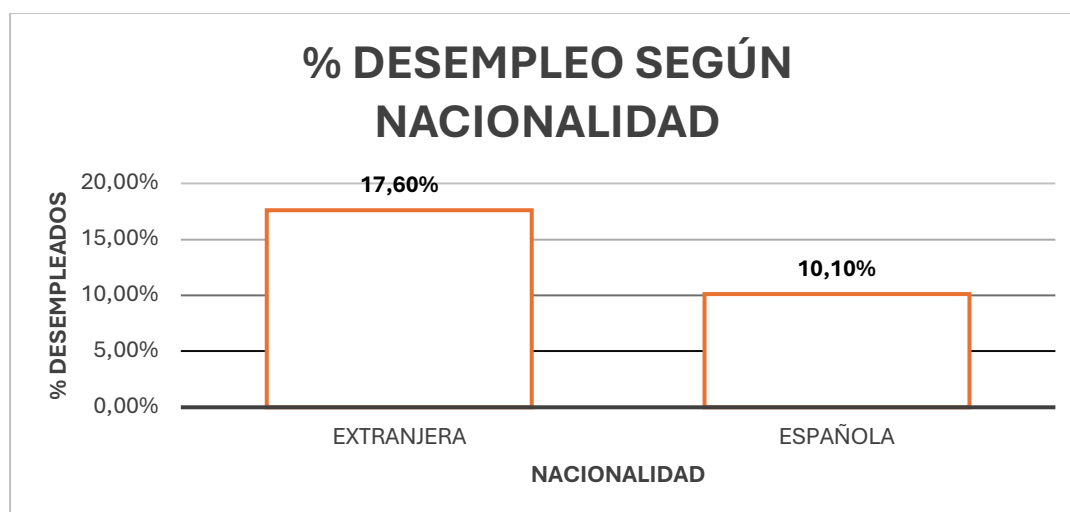
Y, por último, agrupados en estudios superiores, se encuentran los individuos que poseen un grado superior finalizado, son graduados universitarios o estudios superiores, como pueden ser estudios de master o doctorado, para ellos la tasa de desempleo se reduce hasta el 6,4%.

Como cabría esperar, se obtuvo que cuanto mayor es la formación finalizada por los individuos, menor es porcentaje de individuos que se encuentran en una situación de desempleo.

2.2.6. Análisis de la variable NACIONALIDAD

Por último, se analizó la variable NACIONALIDAD para estudiar si existen diferencias determinadas por el territorio de la península en el que tienen su residencia los individuos.

Ilustración 8. Porcentaje (%) de desempleo en función de la variable NACIONALIDAD.



Fuente: Elaboración propia con Excel desde datos EPA

Para analizar esta variable, los individuos se agruparon en función de su nacionalidad, extranjera o española. Se obtuvo que las personas encuestadas con nacionalidad extranjera tienen una alta tasa de desempleo, el 17,6%, mientras que los encuestados con nacionalidad española su tasa de paro desciende hasta el 10,1%, por debajo de la media general de la muestra.

3. MODELO LOGIT

3.1. Marco teórico

El principal objetivo de esta investigación era construir un modelo para tratar de conocer la probabilidad de un individuo de estar en una situación de desempleo, en función de unos determinantes o características del individuo.

La situación de desempleo del individuo se recogió en la variable cualitativa *AOI*, que recoge si el individuo tiene o no trabajo, por lo tanto, nuestra variable dependiente Y_i , es una variable dicotómica con dos posibles valores:

$$Y_i: \begin{cases} 1 & \text{si el individuo está desempleado} \\ 0 & \text{si el individuo está ocupado} \end{cases}$$

Obteniendo las probabilidades de cada valor de la siguiente forma: $P_i = P(Y = 1 | x)$; y en el caso de encontrarse desempleado, $1 - P_i = P(Y = 0 | x)$.

Con estas características, el modelo obtenido se trata de un modelo de elección discreta. Estos modelos tratan de explicar una variable cualitativa (*AOI*) en función de otras variables cualitativas y cuantitativas.

Para nuestro estudio estas otras variables son los determinantes o características de los individuos estudiados, que hemos visto en el análisis descriptivo (*EDAD*, *SEXO*, *ESTADO CIVIL*, *NACIONALIDAD* y *NIVEL DE FORMACIÓN*).

Estas variables independientes son los regresores del modelo, con los que, si suponemos una relación lineal entre estos y la probabilidad de *AOI*, p_i .

$$Y_i = \beta_0 + \beta_1 * EDAD_i + \beta_2 * SEXO_i + \beta_3 * EST.CIVIL_i + \beta_4 * NACIONALIDAD_i + \beta_5 * FORMACIÓN_i^1$$

Por lo que construimos un modelo de regresión lineal múltiple, o MLP².

$$\begin{aligned} E(Y_i | X_{1i}, \dots, X_{ki}) \\ = \beta_0 + \beta_1 * EDAD_i + \beta_2 * SEXO_i + \beta_3 * EST.CIVIL_i + \beta_4 \\ * NACIONALIDAD_i + \beta_5 * FORMACIÓN_i + u_i \end{aligned}$$

Pero este modelo presenta un problema para un estudio con las características del nuestro. Y es que este modelo, al estimarlo y obtener las probabilidades de estar

¹ La variable *FORMACIÓN* agrupa las variables: *PRIMARIA*, *SECUNDARIA*, *BACHILLERATO* y *SUPERIORES*, como los diferentes niveles de estudios que se pueden alcanzar. En la construcción del modelo logit serán estas cuatro variables las usadas como dummy, usando *PRIMARIA* como base.

² MLP o modelo lineal de probabilidad.

desempleado puede dar como resultado tanto valores negativos, como valores superiores a 1, que para nuestro estudio no aportan ninguna conclusión y carecen de sentido.

Para solucionar este inconveniente, utilizamos el modelo logit. La formulación de este modelo permite que las predicciones obtenidas, en nuestro caso probabilidades de estar desempleado, se encuentren en un intervalo entre 0 y 1. Con lo que nuestro modelo tendría la siguiente forma.

$$Y_i = F(\beta_0 + \beta_1 * EDAD_i + \beta_2 * SEXO_i + \beta_3 * EST.CIVIL_i + \beta_4 * NACIONALIDAD_i + \beta_5 * FORMACIÓN_i) + u_i; \quad 0 \leq F(.) \leq 1$$

Donde F es una función logística no lineal, de esperanza 0 y desviación típica $\frac{\pi}{\sqrt{3}}$ con la siguiente forma:

$$F(t) = \frac{e^t}{1 + e^t} = \frac{1}{1 + e^{-t}}$$

Donde esa función t, sería nuestra función lineal; con lo que si sustituimos nuestro modelo de predicción finalmente tendría la siguiente forma:

$$Y_i = \frac{e^{\beta_0 + \beta_1 * EDAD_i + \beta_2 * SEXO_i + \beta_3 * EST.CIVIL_i + \beta_4 * NACIONALIDAD_i + \beta_5 * FORMACIÓN_i}}{1 + e^{\beta_0 + \beta_1 * EDAD_i + \beta_2 * SEXO_i + \beta_3 * EST.CIVIL_i + \beta_4 * NACIONALIDAD_i + \beta_5 * FORMACIÓN_i}}$$

$$= \frac{1}{1 + e^{-(\beta_0 + \beta_1 * EDAD_i + \beta_2 * SEXO_i + \beta_3 * EST.CIVIL_i + \beta_4 * NACIONALIDAD_i + \beta_5 * FORMACIÓN_i)}}$$

Una vez obtenido el modelo, el siguiente paso es interpretar sus resultados, en un modelo lineal el cambio en la probabilidad de que $Y_i = 1$, si se incrementa en una unidad X_{ij} , y el resto permanece constante, es constante:

$$\Delta p = p_i^{+1} - p_i = \beta_j$$

Por lo que la interpretación se puede realizar directamente desde los coeficientes del modelo.

Pero el modelo logit aplicado no es un modelo lineal, lo que implica que los coeficientes dependen del valor de todos los regresores del modelo.

Para interpretarlos es necesario recurrir a los ODD ratio. Un ODD es el cociente entre la probabilidad de que $Y_i = 1$ y la probabilidad de que $Y_i = 0$.

$$\begin{aligned}
 ODD &= \frac{p(Y_i = 1)}{p(Y_i = 0)} = \frac{\frac{e^{X_i' \beta}}{1 + e^{X_i' \beta}}}{\frac{1}{1 + e^{X_i' \beta}}} = e^{X_i' \beta} \\
 &= e^{\beta_0 + \beta_1 * EDAD_i + \beta_2 * SEXO_i + \beta_3 * EST.CIVIL_i + \beta_4 * NACIONALIDAD_i + \beta_5 * FORMACIÓN_i}
 \end{aligned}$$

Y un ODD ratio es el cociente entre dos ODD, por lo que podemos usarlo para medir cuanto cambia un ODD cuando se incrementa en una unidad uno de sus regresores:

$$ODD \text{ ratio} = \frac{e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_j (X_{ij} + 1) + \dots + \beta_k X_{ki}}}{e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_j X_{ij} + \dots + \beta_k X_{ki}}} = e^{\beta_j}$$

Interpretamos ese e^{β_j} respecto de la unidad, de forma que si es < 1 , y el resto de los regresores permanece constante, nos dice que el ODD se reduce. Para un ODD ratio de 0,93 la relación entre la probabilidad de que $Y_i = 1$ y la probabilidad de que $Y_i = 0$ disminuye un 3 %. Y si es > 1 el ODD aumenta, para un ODD ratio de 1.09 la relación entre la probabilidad de que $Y_i = 1$ y la probabilidad de que $Y_i = 0$ aumenta un 9%.

3.2. Análisis del modelo logit

Para realizar la construcción del modelo logit se contó con las variables descritas en el apartado anterior, AOI, EDAD, SEXO, ESTCIVIL, NACIONALIDAD y FORMACION. Todas ellas variables dicotómicas que toman valores de 1 o 0 para sus categorías, a excepción de la variable FORMACION que está formada por cuatro categorías diferentes.

El incluir esta variable el modelo se realizó incluyendo variables dummy en el modelo. La cantidad de variables a incluir se realizó mediante la regla de $m - 1$, siendo m el número de categorías de la variable original. Tras recodificar la variable FORMACION, se obtuvieron las variables dummy: PRIMARIA, SECUNDARIA, BACHILLERATO Y SUPERIORES, correspondientes a cada uno de los diferentes niveles de estudios. Para

el modelo se utilizó como base la variable *PRIMARIA* siendo las otras tres las incluidas en el modelo logit.

Por lo que se plantea el modelo con siguiente forma:

$$AOI_i = F(\beta_0 + \beta_1 EDAD_i + \beta_2 SEXO_i + \beta_3 EST.CIVIL_i + \beta_4 NACIONALIDAD_i + \beta_5 SECUNDARIA_i + \beta_6 BACHILLERATO_i + \beta_7 SUPERIORES_i) + u_i$$

Siendo F la función de distribución logística clásica.

Tabla 2. Estimación del modelo logarítmico

Variables en la ecuación		B	Error estándar	Wald	gl	Sig.	Exp(B)
Paso 1 ^a	EDAD	-,021	,001	307,088	1	<,001	,979
	SEXO	-,449	,028	254,562	1	<,001	,638
	ESTCIVIL	,479	,031	237,983	1	<,001	1,615
	NACIONALIDAD	-,301	,041	53,101	1	<,001	,740
	SECUNDARIA	-,555	,055	102,562	1	<,001	,574
	BACHILLERATO	-1,008	,057	311,959	1	<,001	,365
	SUPERIORES	-1,668	,057	847,234	1	<,001	,189
	Constante	,042	,083	,259	1	,611	1,043

a. Variables especificadas en el paso 1: EDAD, SEXO, ESTCIVIL, NACIONALIDAD, SECUNDARIA, BACHILLERATO, SUPERIORES.

Fuente: Elaboración propia con SPSS desde datos EPA

Esta tabla nos ofrece los resultados de la estimación del modelo logit. En primer lugar, tenemos la columna de los coeficientes de las variables “B”, que al tratarse de un modelo logit únicamente podemos tener en cuenta el signo. Observamos que todas las variables del modelo, excepto *ESTCIVIL*, tienen signo negativo, es decir, que cuanto mayor sea el valor de esas variables menores serán las probabilidades de estar desempleado del individuo. El caso de la variable *ESTCIVIL*, que tiene signo positivo, lo que nos indica, junto con su codificación que, los individuos solteros van a tener una mayor probabilidad de encontrarse en situación de desempleo.

La tabla también nos ofrece tanto el estadístico del Test de Wald como la significación de cada variable, analizándolos en conjunto tenemos que todas las variables introducidas

son significativas, a excepción del término constante, que en la práctica no es relevante para la interpretación del modelo.

Y por el ultimo la columna “Exp(B)”, que nos ofrece los ODD ratio de cada variable. Podemos observar que la variación de los ODDs concuerda con el signo de los coeficientes beta, siendo la variable nacionalidad la única que obtiene un ODD ratio superior a uno, 1,615. Es decir, cuando un individuo cambie su estado civil, de casado a soltero, su ODD, el cociente entre $p(Y_i = 1)$ e $p(Y_i = 0)$, aumentará en 61,5 puntos porcentuales.

Tabla 3. Clasificación del modelo logarítmico

Tabla de clasificación^a

		Pronosticado		Porcentaje correcto
Observado		AOI	AOI	
		OCUPADOS	PARADOS	
Paso 1	AOI			
	OCUPADOS	32083	18970	62,8
	PARADOS	2106	4062	65,9
	Porcentaje global			63,2

a. El valor de corte es ,108

Fuente: Elaboración propia con SPSS desde datos EPA

La tabla de clasificación del modelo nos ofrece una comparativa de como de bien clasifica el modelo que hemos generado a los individuos. Con los porcentajes que nos ofrece podemos decir que el modelo clasifica correctamente en torno a 2 de cada 3 individuos, un 62,8% de los casos para los ocupados. Y en el caso de los parados un 65,9%. Al realizar el modelo hay que tener en cuenta la necesidad de cambiar el valor de corte de la clasificación de 0,5, el valor por defecto, a en nuestro caso 0,108, un 10,8% que es la tasa de desempleados media con la que estamos trabajando. De no hacerlo el modelo clasificaría de forma errónea, a la mayoría de los individuos como ocupados.

3.2.1. Efecto marginal de cada variable

El siguiente paso fue calcular el efecto que cada variable tiene sobre el modelo, es decir, el efecto marginal. En este modelo tenemos que diferenciar entre dos tipos de variables,

continuas como es el caso de la variable *EDAD*, y dicotómicas en el caso del resto de variables. Ya que el efecto marginal se calcula de formas diferentes.

Para calcular el efecto de la variable edad lo haremos mediante la siguiente formula: $P_i * (1 - P_i) * \beta_{EDAD}$, siendo P_i la probabilidad pronosticada de estar en paro o no de cada individuo que nos da SPSS al realizar el modelo logit. Y β_{EDAD} el valor de la variable *EDAD* en el modelo, que en nuestro caso es de -0,021.

Una vez obtenido el efecto marginal de cada uno de los individuos, mediante un análisis descriptivo se obtuvo el efecto marginal medio.

Tabla 4. Efecto marginal de la variable *EDAD*

Estadísticos descriptivos					
	N	Mínimo	Máximo	Media	Desv. estándar
marginalEDAD	57221	,00046	,00525	,0019222	,00099222
N válido (por lista)	57221				

Fuente: Elaboración propia con SPSS desde datos EPA

Como resultado se obtuvo que cuando la edad se incrementa de forma marginal, la probabilidad de estar situación de desempleo disminuye, de media, 0,00192 puntos porcentuales. Hay que tener en cuenta que disminuye ya que el valor de β_{EDAD} es negativo.

Por otro lado, para calcular el efecto marginal de las variables dicotómicas el método cambia, ya que son variables que no pueden aumentar de forma progresiva su valor una unidad, sino que valen 1 o 0, por ejemplo, para el caso de la nacionalidad, el individuo únicamente puede ser español/a o extranjero/a. Por lo tanto, se calcula el valor de la probabilidad cuando la variable en cuestión vale 1 y cuando tiene el valor 0, manteniendo el resto de las variables constantes, en este caso con su valor medio, y su diferencia es el efecto marginal. A excepción del caso de la variable formación, que al tratarse de variables dummy las incluidas en el modelo, *SECUNDARIA*, *BACHILLERATO* y *ESTUDIOS SUPERIORES*, la variable en cuestión tendrá el valor 1 y las demás el valor 0 en lugar del valor medio como en el resto de los casos.

En primer lugar, estudiamos el efecto de las variables dummy correspondientes al nivel de formación, recordando que la variable *PRIMARIA* ejerce de base.

Tabla 5. Efecto marginal de la variable FORMACIÓN

	$p(Y_i = 1 X_i = 1)$	$p(Y_i = 1 X_i = 0)$	Efecto marginal $p(Y_i = 1 X_i = 1) - p(Y_i = 1 X_i = 0)$
SECUNDARIA	0,1558	0,2433	- 0,0875
BACHILLERATO	0,1050	0,2433	- 0,1383
SUPERIORES	0,0572	0,2433	- 0,1861

Fuente: Elaboración propia

Esto nos indica que, manteniendo el resto de las variables constantes en su valor medio, y teniendo como base los estudios primarios, el tener unos estudios secundarios aporta un 15,58% de probabilidades de que el individuo se encuentre en situación de desempleo, lo que supone una reducción de 8,75 puntos porcentuales respecto a primaria. Si en lugar de tener unos estudios primarios, por el contrario, el individuo finalizó el bachillerato o equivalente, la aportación del nivel de estudios se reduce hasta un 10,50, es decir, una diferencia de 13,83 puntos porcentuales entre primaria y bachillerato. Y en el caso de tener finalizados estudios superiores la probabilidad de estar en situación de desempleo asociada a este nivel de estudios baja hasta un 5,72%, lo que marca una diferencia de 18,61 puntos porcentuales entre la educación primaria y la superior.

Tabla 6. Efecto marginal de las variables SEXO, ESTCIVIL y NACIONALIDAD

	$p(Y_i = 1 X_i = 1)$	$p(Y_i = 1 X_i = 0)$	Efecto marginal $p(Y_i = 1 X_i = 1) - p(Y_i = 1 X_i = 0)$
SEXO	0,2051	0,2879	- 0,0828
ESTCIVIL	0,2920	0,2035	0,0885
NACIONALIDAD	0,2383	0,2971	- 0,0588

Fuente: Elaboración propia

En el caso de las demás variables nos encontramos que, para el sexo, es un 8,28% más probable encontrarse en paro siendo mujer que siendo hombre. Para el estado civil, los individuos casados tienen una menor probabilidad de estar desempleados con una diferencia de 8,85 puntos porcentuales, 29,20% frente a 20,35%. Y, por último, en el caso de la nacionalidad los individuos con nacionalidad extranjera tienen una mayor probabilidad de estar en situación de desempleo, con un 29,71% frente al 23,83% de los individuos con nacionalidad española, una diferencia de 5,88 puntos porcentuales.

4. CREACIÓN Y ANÁLISIS DE LOS ESCENARIOS

Una vez construido el modelo logit, y estudiados los efectos marginales de cada una de las variables, pasamos a construir una serie de escenarios, con el objetivo de conocer cuál es la probabilidad que le modelo otorga a diferentes combinaciones de características de los individuos.

Para ello hemos creado pequeños grupos, entre tres y cuatro individuos de cada edad, y a partir de esos grupos, estudiar diferentes combinaciones de individuos.

Tabla 7. Escenarios

Edad	Sexo	Estado Civil	Nacionalidad	Nivel Estudios	Probabilidad pronosticada
20	Mujer	Soltera	Extranjera	Primaria	0,5252
20	Hombre	Soltero	Española	Secundaria	0,2308
20	Hombre	Soltero	Extranjera	Bachillerato	0,2049
25	Mujer	Soltera	Española	Superiores	0,1221
25	Hombre	Casado	Española	Primaria	0,2257
25	Mujer	Casada	Extranjera	Primaria	0,3815
30	Mujer	Casada	Extranjera	Secundaria	0,2418
30	Hombre	Casado	Española	Bachillerato	0,0874
30	Mujer	Soltera	Extranjera	Superiores	0,1447

35	Mujer	Soltera	Española	Bachillerato	0,1790
35	Hombre	Casado	Española	Secundaria	0,1194
35	Mujer	Casada	Española	Primaria	0,2701
35	Hombre	Soltero	Extranjera	Bachillerato	0,1583
40	Mujer	Soltera	Española	Superiores	0,0921
40	Hombre	Casado	Española	Primaria	0,2556
40	Mujer	Casada	Extranjera	Secundaria	0,2054
45	Hombre	Soltero	Extranjera	Superiores	0,0730
45	Mujer	Casada	Española	Bachillerato	0,0987
45	Hombre	Casado	Extranjera	Bachillerato	0,0863
50	Hombre	Casado	Española	Secundaria	0,0901
50	Mujer	Soltera	Extranjera	Primaria	0,3708
50	Hombre	Soltero	Española	Superiores	0,0499
55	Hombre	Casado	Española	Bachillerato	0,0536
55	Mujer	Casada	Española	Superiores	0,0439
55	Hombre	Soltero	Extranjera	Primaria	0,2529
60	Mujer	Soltera	Española	Secundaria	0,1687
60	Hombre	Casado	Española	Superiores	0,0257
60	Mujer	Casada	Española	Primaria	0,1796
65	Mujer	Casada	Extranjera	Bachillerato	0,0478
65	Hombre	Soltero	Española	Primaria	0,1688

Fuente: Elaboración propia

Los primeros tres escenarios se tratan de individuos de 20 años. Se observan probabilidades de encontrarse en situación de desempleo bastante elevadas, es especialmente relevante el caso de una mujer soltera, con nacionalidad extranjera y

estudios primarios que refleja una probabilidad del 52,52%, la más alta de todos los escenarios. En el caso de los hombres las probabilidades bajan, para un hombre soltero, con nacionalidad española y estudios secundarios la probabilidad de estar desempleado baja hasta un 23,08%; y para ese mismo individuo, pero con nacionalidad extranjera y estudios de bachillerato la probabilidad continúa bajando hasta el 20,49%.

Esto nos hace ver la importancia del nivel de educación, especialmente el salto de la educación primaria a la secundaria, incluso entre los individuos extranjeros.

En el grupo de 25 años tenemos tres individuos, dos mujeres y un hombre. Nos encontramos con una mujer soltera, con nacionalidad española y con estudios superiores que tiene una probabilidad de estar desempleada del 12,21%, se trata ya de una probabilidad cercana a la media 10,8%, en la cual el motivo de encontrarse ligeramente por encima se deba a la corta edad del individuo que implicará seguramente una baja experiencia laboral.

En los otros dos casos nos permite comparar la aportación del sexo y la nacionalidad, ya que se trata de un hombre y una mujer, ambos casados y con estudios primarios, teniendo el hombre nacionalidad española y la mujer nacionalidad extranjera, con un pronóstico de un 22,57% para el hombre y un 38,15% para la mujer.

A continuación, en el grupo de 30 años nos volvemos a encontrar con tres individuos. Que nos permiten hacer hincapié en la importancia del nivel de estudios. Observamos una mujer casada con nacionalidad extranjera y estudios de secundaria y otra mujer, soltera, también con nacionalidad extranjera, pero con estudios superiores, la diferencia en las probabilidades de estar en situación de desempleo obtenidas es relevante, un 24,18% y un 14,47% respectivamente.

Destaca el tercer caso, un hombre casado, con nacionalidad española y estudios de bachillerato, que tiene un 8,74% de probabilidad, siendo el primer individuo estudiado en encontrarse por debajo de la media.

Continuamos con cuatro individuos de 35 años, en este caso se trata de dos hombres y dos mujeres. En el caso de las mujeres vemos que ambas tienen nacionalidad española,

y una de ellas está soltera, con estudios de bachillerato, que hace que tenga una probabilidad pronosticada del 17,9%; y la segunda mujer está casada y tiene estudios primarios, lo que resulta en una probabilidad del 27,01%. De nuevo volvemos a ver el gran peso que tiene el nivel de estudios, que elevan la probabilidad de una situación de desempleo especialmente en el caso de los estudios primarios.

En el caso de los dos hombres, el primero está casado, con nacionalidad española y estudios secundarios, y una probabilidad de desempleo del 11,94%; mientras que el segundo hombre, soltero con nacionalidad extranjera y estudios de bachillerato que tiene una probabilidad de desempleo del 15,83%. En esta comparación entre los dos individuos varones, resulta interesante observar como la relación entre el estado civil y la nacionalidad del primero, casada y española, tiene un mayor peso a la hora de disminuir la probabilidad de desempleo, que un nivel de estudios superior en el caso del segundo hombre.

Continuamos con los individuos de 40 años, en este caso tres. De nuevo en este grupo se ve la importancia de un nivel de estudios elevado para una probabilidad de desempleo lo más baja posible. Observamos una mujer soltera, con nacionalidad española y estudios superiores que tiene una probabilidad pronosticada del 9,21%, por debajo de la media. Mientras que, por otro lado, tenemos un hombre y una mujer, ambos casados, el hombre con nacionalidad española y estudios primarios y una probabilidad de desempleo del 25,56%; mientras que la mujer con nacionalidad extranjera y estudios secundarios tiene una probabilidad del 20,54% de estar desempleada. De nuevo vemos que el salto de estudios primarios a secundarios resulta básico para reducir la probabilidad de desempleo. En este grupo pese a unos niveles de estudios reducidos, se empiezan a ver probabilidades algo menores, seguramente efecto de una mayor edad lo que puede implicar una mayor experiencia laboral.

Para el grupo de edad de 45 años tenemos de nuevo tres individuos. Una mujer casada, con nacionalidad española y estudios de bachillerato, con una probabilidad pronosticada del 9,87% que se sitúa por debajo de la media, este escenario puede ser muy representativo en la población en España.

Y dos hombres, ambos con nacionalidad extranjera, uno de ellos soltero y con estudios superiores, y el segundo, casado y con estudios de bachillerato, ambos tienen probabilidades pronosticadas bajas, por debajo de la media, un 7,30% y un 8,63% respectivamente. Son probabilidades de desempleo muy parejas, lo que nos indica que el estado civil no tiene un gran peso en estos casos y que el salto en el nivel de estudios de bachillerato a estudios superiores genera diferencias, pero mínimas.

A continuación, el grupo de los individuos con 50 años cuenta con 3 individuos. Una mujer soltera, con nacionalidad extranjera y estudios primarios, que resulta en una probabilidad de desempleo del 37,08%, esta combinación salvo por el factor experiencia laboral relacionado con la edad es una de las que otorga una mayor probabilidad de encontrarse sin trabajo.

Y dos hombres, ambos con nacionalidad española, uno casado con estudios secundarios que tiene una probabilidad pronosticada del 9,01%; y el segundo, soltero con estudios superiores y con una probabilidad del 4,99%. De nuevo, pese a encontrarnos en un rango de edad más avanzado el nivel de estudios tiene una gran relevancia a la hora de obtener un mejor pronóstico.

Para el grupo de 55 años tenemos tres individuos. Destaca el caso de un hombre soltero, con nacionalidad extranjera y estudios primarios, que tiene un pronóstico de estar desempleado del 25,29%, de nuevo una probabilidad muy elevada condicionada por la combinación dada principalmente en la nacionalidad y el nivel de estudios.

Y un hombre y una mujer, ambos casados y con nacionalidad española, con estudios de bachillerato el hombre y una probabilidad de desempleo del 5,36%; y la mujer con estudios superiores, y una probabilidad pronosticada del 4,39%. Estos dos escenarios en conjunto podrían ser la representación en la sociedad española el caso de unos padres de familia.

Para el grupo de 60 años, contamos con otros tres individuos. En primer lugar, una mujer soltera, con nacionalidad española y estudios secundarios que tiene una probabilidad del 16,87% de desempleo.

Y un hombre y una mujer, de nuevo ambos casados y con nacionalidad española, esta vez la mujer con estudios primarios resulta en una probabilidad pronosticada del 17,96%. Y el hombre con estudios superiores tiene una probabilidad de desempleo del 2,57%, que se trata de la más baja obtenida en todos los escenarios.

Este dato viene condicionado por la combinación de una edad ya avanzada, lo que otorga una gran experiencia laboral, y el alto nivel de estudios.

Por último, el grupo de los 65 años cuenta con dos individuos. Una mujer casada, con nacionalidad extranjera y estudios de bachillerato con una baja probabilidad de desempleo, un 4,78%.

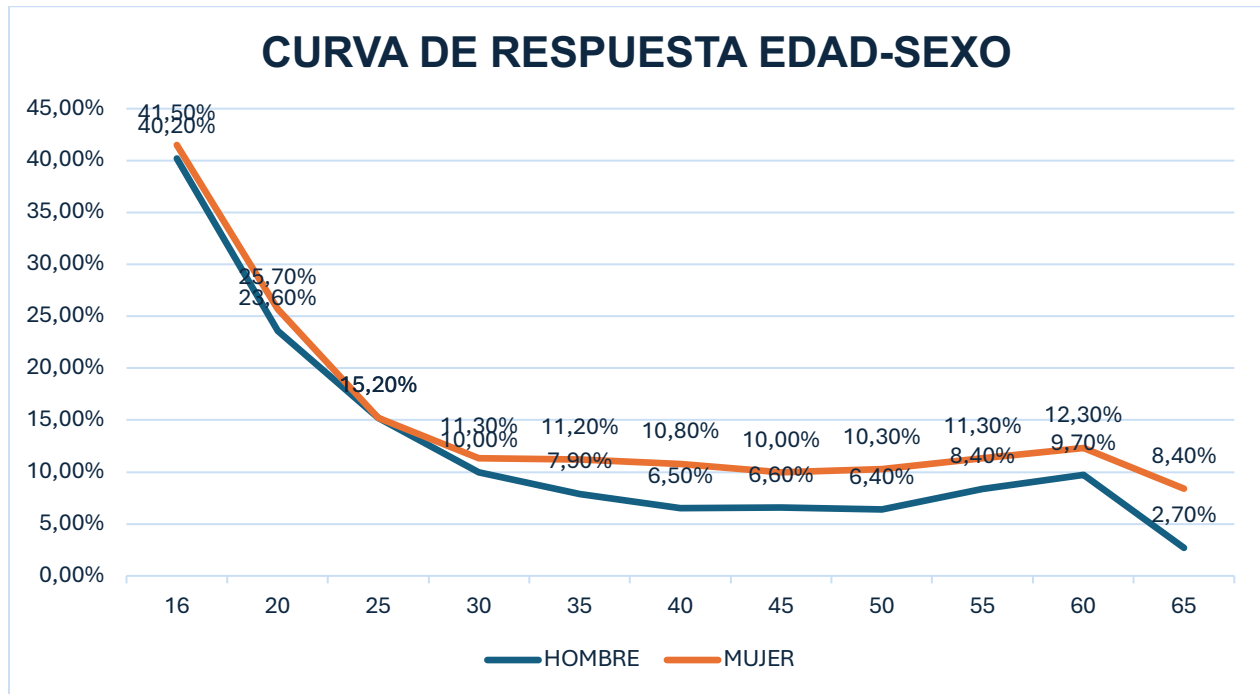
Y un hombre soltero, con nacionalidad española, y estudios primarios, que resulta en una probabilidad de desempleo del 16,88%.

4.1. Curvas de respuesta

Por último, vamos a analizar la evolución de la tasa de paro, del periodo que escogimos al inicio de la investigación, para cada una de las diferentes variables *SEXO*, *NACIONALIDAD*, *ESTCIVIL* y *FORMACIÓN*, a lo largo de las diferentes edades de los individuos.

Con el objetivo de analizar de una forma más visual como va cambiando esa tasa de desempleo.

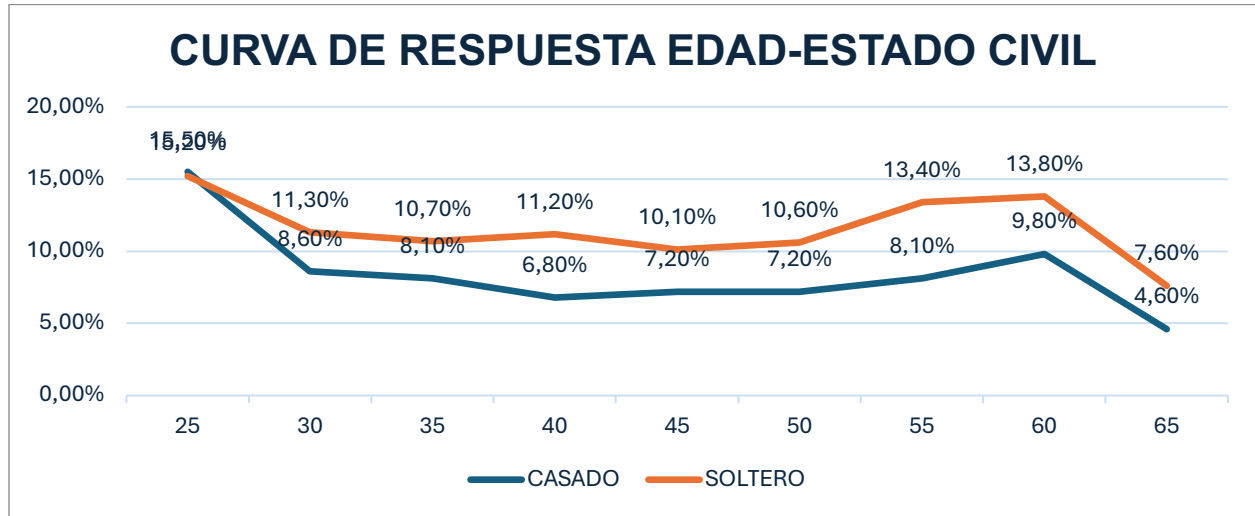
Ilustración 9. Gráfico curva EDAD-SEXO



Fuente: Elaboración propia con Excel desde datos EPA

Este gráfico representa la comparación de la probabilidad de estar en una situación de desempleo a lo largo de diferentes edades para hombres y mujeres. Se observa que las mujeres siempre tienen una probabilidad más elevada que los hombres, entre los 16 y los 30 años estas diferencias son mínimas, incluso en el caso de los 25 años, la probabilidad se iguala para ambos sexos, pero a partir de los 30 hasta los 65 años, estas diferencias son algo mayores, de entre un 3% y un 4%.

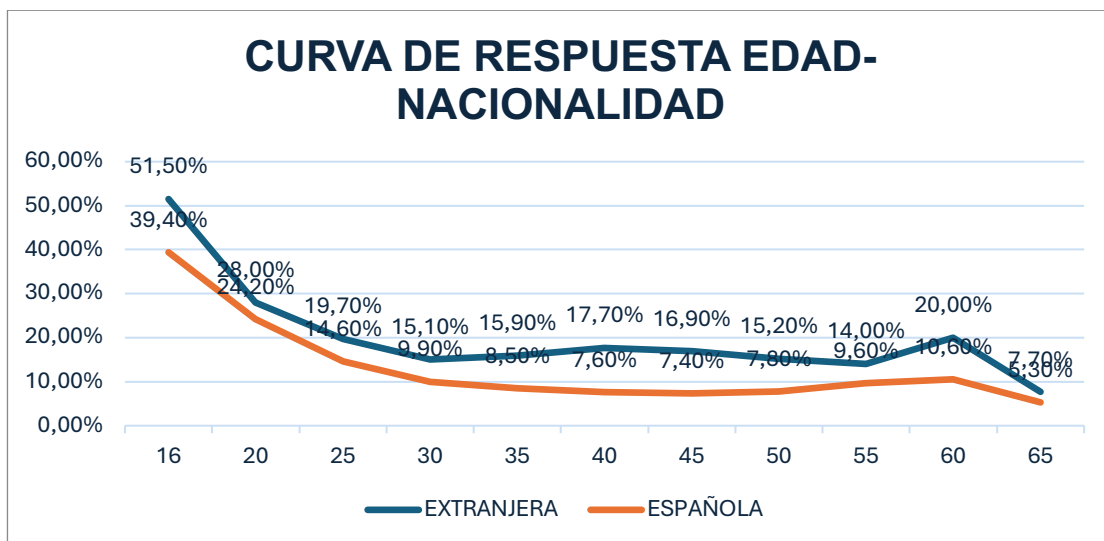
Ilustración 10. Gráfico curva EDAD-ESTADO CIVIL



Fuente: Elaboración propia con Excel desde datos EPA

En este caso la comparación se realiza con el estado civil de los individuos. Se observa que, en edades tempranas, hasta los 25 años, la probabilidad de estar en paro de los casados supera a la de los solteros, seguramente debido a los pocos individuos no solteros a tan temprana edad. Es a partir de los 25 años, cuando la probabilidad de las personas solteras supera, por un porcentaje pequeño, a la de los casados.

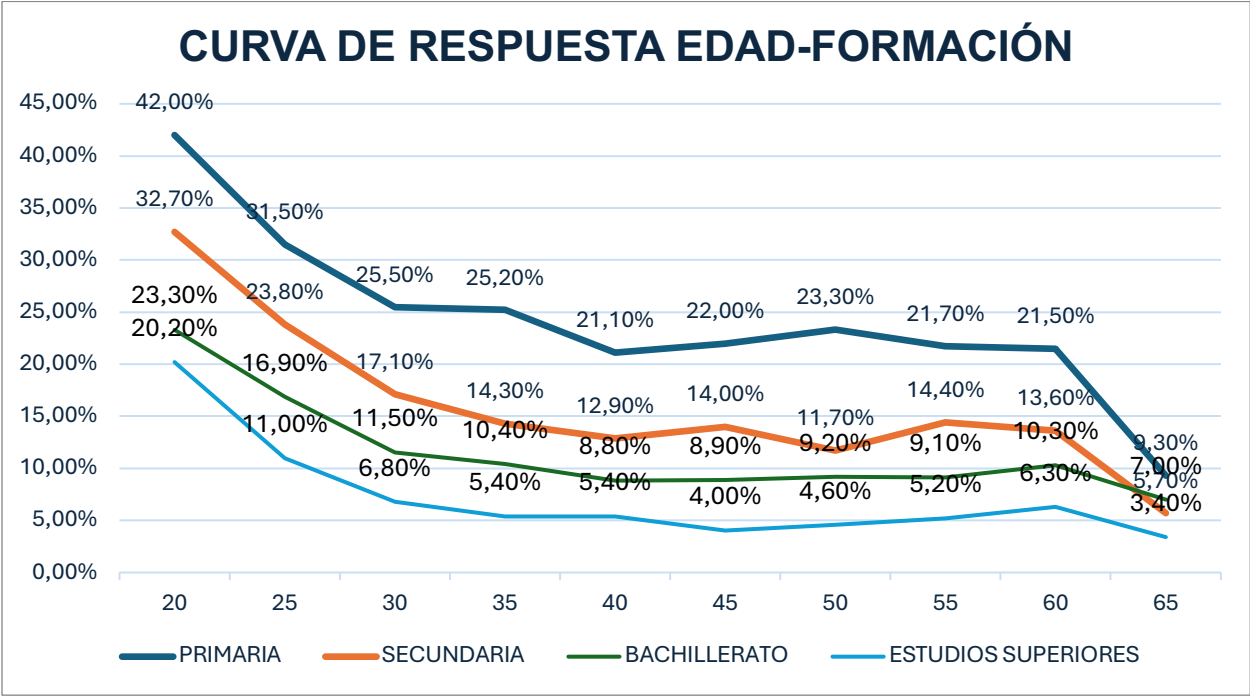
Ilustración 11. Gráfico curva EDAD-NACIONALIDAD



Fuente: Elaboración propia con Excel desde datos EPA

Para el caso de la nacionalidad encontramos que la probabilidad de encontrarse desempleado es siempre mayor para los individuos con nacionalidad extranjera, que, para aquellos con nacionalidad española, independientemente de la edad. Esta diferencia varía, llegando en algunos tramos de edad a valores superiores al 10%.

Ilustración 12. Gráfico curva EDAD-FORMACIÓN



Fuente: Elaboración propia con Excel desde datos EPA

Por último, analizamos el caso del nivel de formación. Se observa claramente que el desempleo disminuye conforme aumenta la formación: las personas con estudios superiores presentan en todo momento las tasas más bajas de desempleo, seguidas de quienes tienen bachillerato, secundaria y, por último, primaria, que registra los porcentajes más altos.

Además, vemos que las diferencias entre niveles educativos tienden a reducirse al acercarse a los 65 años, momento de la edad habitual de jubilación, donde las probabilidades de desempleo se aproximan en todos los grupos. Esto puede deberse a

que muchas personas abandonan el mercado laboral o acceden a jubilaciones anticipadas, igualando así la situación laboral independientemente del nivel de estudios.

5. CONCLUSIONES

A lo largo de este estudio hemos analizado los principales determinantes del desempleo en España, centrándonos en variables relacionadas con el ámbito personal como el sexo, la edad, el estado civil, la nacionalidad y el nivel de formación. El análisis descriptivo inicial permitió identificar algunas tendencias claras. Si bien en edades jóvenes las diferencias entre hombres y mujeres son relativamente pequeñas, a medida que avanza la edad se observa que las mujeres presentan una mayor probabilidad de encontrarse en situación de desempleo que los hombres.

En lo referente a la edad, se observa que la presencia del denominado paro juvenil, donde los individuos más jóvenes son los que mayores dificultades tienen para acceder al mercado laboral. Esta situación se debe principalmente a la falta de experiencia profesional y, en muchos casos, al hecho de no haber completado aún estudios superiores.

El estado civil también muestra diferencias notables. Los resultados indican que las personas casadas tienen una menor probabilidad de desempleo frente a las solteras. Esto puede estar relacionado con la edad, ya que los individuos jóvenes, que son quienes más sufren el desempleo, suelen ser mayoritariamente solteros.

Por su parte, la nacionalidad constituye otro factor relevante: las personas extranjeras encuentran más obstáculos para incorporarse al mercado laboral en comparación con las personas de nacionalidad española. No obstante, este efecto podría estar relacionado con el nivel de formación, ya sea por un menor nivel de formación de las personas extranjeras, o como sucede en algunos casos que extranjeros con estudios superiores encuentran dificultades para que se les reconozcan sus estudios.

Entre todos los factores analizados, el nivel de formación destaca como el más determinante. El estudio muestra una relación inversa clara: a mayor nivel educativo, menor probabilidad de encontrarse en situación de desempleo. Esta influencia supera incluso a la de otras variables como el estado civil o la nacionalidad.

El análisis territorial, a través de la variable de comunidad autónoma, mostró ciertas diferencias en el nivel de desempleo entre regiones, aunque no arrojó conclusiones suficientemente sólidas para incorporarlas en el modelo logit final.

Mediante la estimación de un modelo logit, pudimos cuantificar la contribución de cada variable y calcular la probabilidad de desempleo individual en función de diferentes combinaciones de características personales. Los resultados del modelo confirmaron lo observado previamente en los análisis descriptivos. Así, al simular diversos escenarios, se concluye que el perfil con menor probabilidad de desempleo corresponde a un hombre, de edad avanzada, casado, de nacionalidad española y con estudios superiores. En el extremo opuesto, el escenario con mayor riesgo de desempleo es el de una mujer joven, soltera, extranjera y con estudios primarios.

Cabe destacar que, aunque el nivel de formación es el principal determinante, ciertas combinaciones de nacionalidad, estado civil y sexo pueden generar probabilidades de desempleo superiores o inferiores a las que cabría esperar solo por el nivel educativo. Esto indica la importancia de analizar el fenómeno desde una perspectiva global.

En definitiva, pese a la mejora registrada en los últimos periodos en las tasas globales de desempleo, los resultados de este estudio confirman la existencia de un problema estructural en España, especialmente vinculado al desempleo juvenil y a las personas con menor cualificación. Además, se evidencia que la probabilidad de estar empleado o desempleado continúa dependiendo, en gran medida, del nivel de formación alcanzado, lo que sugiere una fuerte asociación entre la titulación académica y el acceso al empleo, que limita las oportunidades de aquellos colectivos con menores recursos educativos.

6. BIBLIOGRAFÍA

Banco de España. (2025). *Boletín económico*.

<https://www.bde.es/bde/es/secciones/informes/boletin-economico/>

Goldberger, A. S. (2001). *Introducción a la econometría*. Ariel.

Instituto Nacional de Estadística (INE). (2025). *Encuesta de Población Activa (EPA), primer trimestre de 2025*.

https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176918&menu=resultados&idp=1254735976595

OCDE. (2025). *Perspectivas económicas de la OCDE: España*.

<https://www.oecd.org/economy/outlook/>

Stock, J. H., & Watson, M. W. (2012). *Introducción a la econometría* (3ª ed., pp. 275–300). Pearson Educación.

Wooldridge, J. M. (2001). *Introducción a la econometría: Un enfoque moderno* (1ª ed., pp. 529–540). International Thomson Editores.