



Universidad de Valladolid
Facultad de Ciencias Económicas
y Empresariales

Trabajo de Fin de Grado

**Grado en Administración y Dirección de
Empresas**

**UN MODELO LOGIT PARA EL ANÁLISIS DE
LOS DETERMINANTES DEL DESEMPLEO**

Presentado por:

Miguel de Blas Fernández

Valladolid, 05 de julio de 2023

Resumen

El desempleo es y ha sido uno de los principales problemas en nuestro país. En este trabajo vamos a tratar de analizar cómo afectan determinadas características personales en el desempleo, para ello, vamos a centrarnos en una muestra de individuos obtenida a través de la Encuesta de Población Activa, recopilada por el Instituto Nacional de Estadística. El desarrollo consistirá en un análisis descriptivo de las variables edad, sexo, estado civil, nacionalidad y nivel de estudios. Posteriormente, utilizaremos un modelo logit para estimar una serie de probabilidades de estar desempleado dependiendo de las características que presente un individuo.

Los resultados del modelo reflejan que las variables con más incidencia a la hora de encontrar empleo son tanto el nivel de estudios que haya alcanzado el individuo como el estado civil en el que se encuentre.

Abstract

Unemployment is and has been one of the main problems in our country. In this paper we will try to analyse how certain personal characteristics affect unemployment by focusing on a sample of individuals obtained from the Labour Force Survey, compiled by the National Institute of Statistics. The development will consist of a descriptive analysis of the variables age, sex, marital status, nationality and level of studies. Subsequently, we will use a logit model to estimate a series of probabilities of being unemployed depending on the characteristics that an individual presents.

The results of the model show that the variables with the greatest impact on finding employment are both the level of studies that the individual has attained and the marital status in which he or she finds himself or herself.

Palabras clave: desempleo, análisis descriptivo, modelo logit, probabilidad.

Códigos JEL: C01, C35, J64.

ÍNDICE

1. INTRODUCCIÓN	1
2. EL DESEMPLEO EN ESPAÑA	3
3. METODOLOGÍA	5
4. ANÁLISIS DESCRIPTIVO	6
4.1 Obtención y tratamiento de los datos	6
4.2 Recodificación de las variables	7
4.3 Análisis de las variables	9
4.3.1 Tasa de desempleo por edades	9
4.3.2 Tasa de desempleo según el sexo.....	10
4.3.3 Tasa de desempleo según la nacionalidad	12
4.3.4 Tasa de desempleo según el nivel de formación	13
4.3.5 Tasa de desempleo según el estado civil.....	15
5. MODELO LOGIT	18
5.1 Análisis del modelo	22
5.1.1 Efectos parciales medios de cada variable	25
5.1.2 Creación de escenarios	28
5.1.3 Curvas de respuesta.....	31
6. CONCLUSIONES	34
7. REFERENCIAS BIBLIOGRÁFICAS	37
8. ANEXOS	39

ÍNDICE DE TABLAS Y GRÁFICOS

Tabla 4.1 Resumen estadístico de la variable sexo.....	10
Tabla 4.2 Resumen estadístico de la variable nacionalidad	12
Tabla 4.3 Resumen estadístico de la variable nivel de formación	13
Tabla 4.4 Resumen estadístico de las variables nacionalidad y nivel de formación.....	14
Tabla 4.5 Resumen estadístico de la variable estado civil	15
Tabla 4.6 Resumen estadístico de las variables estado civil y sexo	17
Tabla 5.1 Estimación del modelo logit.....	23
Tabla 5.2 Odds ratio	24
Tabla 5.3 Efectos medios al cambiar de categoría en las variables cualitativas.....	27
Gráfico 3.1 Evolución del desempleo en España	3
Gráfico 3.2 Comparación de las tasas de paro históricas entre España y la Unión Europea.....	4
Gráfico 4.1 Tasa de desempleo según la variable edad.....	9
Gráfico 4.2 Tasa de desempleo según la variable sexo	11
Gráfico 4.3 Tasa de desempleo según la variable nacionalidad	12
Gráfico 4.4 Tasa de desempleo según la variable nivel de formación.....	14
Gráfico 4.5 Tasa de desempleo según las variables nacionalidad y nivel de formación	15
Gráfico 4.6 Tasa de desempleo según la variable estado civil.....	16
Gráfico 4.7 Tasa de desempleo según las variables estado civil y sexo...	18
Gráfico 5.1 Curva de respuesta de la variable sexo en función de la variable edad	31
Gráfico 5.2 Curva de respuesta de la variable nacionalidad en función de la variable edad	32
Gráfico 5.3 Curva de respuesta de la variable estado civil en función de la variable edad	32
Gráfico 5.4 Curva de respuesta según los niveles de formación en función de la variable edad	33

1. INTRODUCCIÓN

El desempleo ha sido objeto de extensos estudios e investigaciones en el ámbito social y económico, se trata de un problema complejo que involucra múltiples factores relacionados entre sí. La urgencia de abordar este problema ha impulsado numerosas teorías que tratan de explicar por qué sucede este fenómeno tan presente en las economías actuales. Según Keynes, para alcanzar la situación de pleno empleo se requiere de buenas actuaciones del gobierno modificando la demanda agregada por vía impuestos o gasto público para generar empleo dentro del mercado laboral. Otros autores de carácter más liberal defienden que el sector institucional no debe intervenir en el libre funcionamiento del mercado laboral (Encarnación Rocío Cordón Lagares, 2012).

La situación de pleno empleo no es más que una situación ficticia, ya que el desempleo es una variable muy difícil de eliminar al completo. En diversos países supone un verdadero reto hacer frente a esta perturbación. La economía española es un claro ejemplo, ya que presenta una tasa de desempleo¹(Ameco, 2022) del 12,9% en el año 2022; si comparamos esta cifra con Alemania, una de las potencias económicas de la Unión Europea, observamos que tiene una tasa de desempleo del 3,1%, lo que representa una diferencia de prácticamente 10 puntos porcentuales. Asimismo, si comparamos con otro país vecino de similares características como es Italia, cuya tasa de paro se encuentra en el 8,1%, sigue existiendo una diferencia en torno a 5 puntos porcentuales. Más adelante en nuestro trabajo vamos a ahondar más en el desempleo en el caso español, pero es evidente que supone uno de los problemas principales para nuestro país.

Sin embargo, el objetivo principal de este estudio no consiste en explicar el desempleo desde una perspectiva económica que trate de buscar soluciones relevantes para reducirlo.

Nos vamos a enfocar en estudiar cómo afectan una serie de características personales en la destrucción de empleo, utilizando una muestra obtenida a través del Instituto Nacional de Estadística.

¹ Cada una de las tasas de paro utilizadas han sido obtenidas a partir de la base de datos de la AMECO (Annual Macro-economic database of the European Commission's Directorate).

Estas características personales se resumen en cuatro variables independientes: la edad, el sexo, el estado civil y el nivel de formación; es importante mencionar las limitaciones que presentan nuestro estudio, ya que no se han incluido variables explicativas relevantes como pueden ser el salario por trabajador o los años de experiencia que presentan los individuos.

Debido a la importancia que supone el desempleo en nuestro país, se ha decidido escoger la probabilidad de estar desempleado como la variable dependiente a analizar dentro de nuestro modelo logístico junto con las variables independientes anteriormente mencionadas; el objetivo de este modelo es conseguir que nos arroje unos pronósticos coherentes acerca de la probabilidad de estar desempleado de un individuo dadas sus características personales.

Para concluir con nuestra introducción, vamos a resumir la estructura del trabajo en las siguientes fases:

- Introducción del desempleo en España haciendo referencia a la estructura económica de nuestro país y comparando las tasas de paro históricas con la Unión Europea.
- Breve descripción acerca del tratamiento de los datos y la utilización de las herramientas informáticas correspondientes para la creación de nuestro modelo.
- Realización de un análisis descriptivo de nuestras variables extraídas de la muestra seleccionada calculando las tasas de desempleo correspondientes.
- Explicación teórica del modelo logit así como su estimación e interpretación de los resultados obtenidos. Vamos a cuantificar, a partir del cálculo de unas probabilidades estimadas, cuáles son las variables más influyentes a la hora de encontrar empleo.
- Conclusiones finales de nuestro estudio del desempleo acordes a los objetivos fijados.

2. EL DESEMPLEO EN ESPAÑA

Antes de comenzar con nuestro estudio descriptivo y probabilístico, vamos a desarrollar nuestra variable más importante a analizar, el desempleo, más concretamente dentro del territorio español.

Gráfico 2.1 Evolución del desempleo en España



Fuente: elaboración propia a partir de datos consultados en la AMECO

Atendiendo al Gráfico 2.1 podemos observar como la tasa de desempleo² histórica en España está marcada por una situación irregular con numerosas oscilaciones a lo largo de los años. Esto no es de extrañar, ya que, en este país, el mercado laboral se ha caracterizado por un modelo económico que se centra en empleos con poca estabilidad y con bajos niveles de cualificación y remuneración (Centro de Investigaciones Sociológicas, 2014).

Adicionalmente, el tejido productivo en España a partir de la etapa de transición democrática ha estado caracterizado por una baja productividad y un escaso valor agregado. Este fenómeno se debe en parte al proceso de terciarización, que ha llevado a que la mayoría de los empleos se concentren en el sector de los servicios (Centro de Investigaciones Sociológicas, 2014).

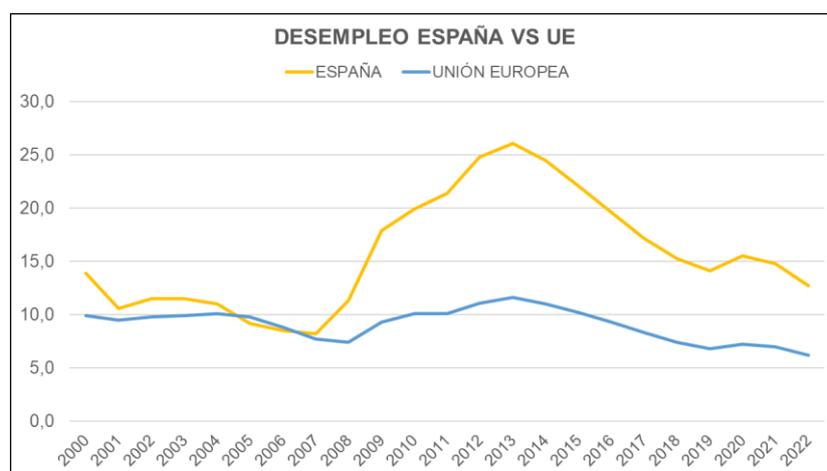
² La tasa de desempleo se refiere a la medida utilizada para calcular el porcentaje de personas en edad de trabajar que se encuentran sin empleo y que están buscando activamente trabajo.

Estas son las principales razones por las cuales el desempleo estructural³ sigue siendo un problema constante en nuestro país. Sin embargo, es importante destacar que se están implementando acciones para tratar de reducirlo. Algunas de estas acciones, entre otras, son:

- Reducción de la excesiva contratación temporal y promover la contratación indefinida.
- Mejoras en la legislación laboral para fomentar la flexibilidad y adaptabilidad del mercado laboral.

Tal y como hemos mencionado anteriormente, España se caracteriza por tener unas elevadas tasas de desempleo en comparación con varios países europeos. Cabe destacar que desde su entrada en la Unión Europea en el año 1986 ha intentado converger con la UE en cuanto a términos de estabilidad económica, política y social.

Gráfico 2.2 Comparación de las tasas de paro históricas entre España y la Unión Europea



Fuente: elaboración propia a partir de datos consultados en la AMECO

Si nos detenemos en el Gráfico 2.2, podemos observar como a principios de la década, la tasa de desempleo entre España y la Unión Europea converge en torno a un 10%.

³ El desempleo estructural hace referencia a la tasa de desempleo presente en el largo plazo y que se origina debido a una inadecuada estructura económica.

Sin embargo, como consecuencia de la caída de Lehman Brothers en 2008, la destrucción de empleo tuvo un crecimiento notorio durante los próximos años y el desempleo cíclico⁴ se disparó hasta alcanzar un máximo histórico en el caso español del 26,1% en comparación al 11,6% que representaban los estados miembros.

Después de esta catástrofe, la Unión Europea tomó medidas para tratar de reducir el desempleo en aquellos países que habían obtenido unas tasas de paro muy elevadas como consecuencia de esta crisis financiera. Es por ello que, según el artículo 148 del TFUE⁵, se exigió la realización de un informe conjunto sobre el empleo anual elaborado por el Consejo y la Comisión. (Versión Consolidada del Tratado de Funcionamiento de la Unión Europea, 2010).

La elaboración del Informe Conjunto sobre Empleo sigue vigente en la actualidad y se recoge en el Calendario del Semestre Europeo, más concretamente dentro del paquete de otoño. Este informe trata de supervisar las medidas tomadas por los países miembros para garantizar la igualdad de oportunidades en el acceso al empleo y para promover mercados laborales justos y dinámicos (Comisión Europea, 2022).

3. METODOLOGÍA

En este apartado vamos a hacer referencia al tratamiento de los datos que hemos tenido que llevar a cabo y la utilización de los distintos programas informáticos.

En primer lugar, se han descargado en un fichero Excel los microdatos correspondientes al cuarto trimestre del año 2022 de la EPA⁶ a través del INE⁷.

⁴ El desempleo cíclico hace referencia a la tasa de paro que se manifiesta en fases de expansión o recesión de la economía.

⁵ Tratado de Funcionamiento de la Unión Europea.

⁶ Encuesta de Población Activa: se realiza a través de una muestra representativa de hogares y se recopila información sobre la situación laboral de las personas mayores de 16 años.

⁷ Instituto Nacional de Estadística: se trata de una organización autónoma del Estado que tiene como función principal producir y difundir información estadística oficial de alta calidad, relevante y objetiva sobre la realidad social, económica y demográfica de España.

En segundo lugar, tras obtener de la muestra aquellas variables que consideramos relevantes, es decir, el desempleo, la edad, el sexo, el estado civil, la nacionalidad y el nivel de formación, exportaremos el fichero de los datos al software Statgraphics para calcular las tasas de desempleo y así proceder con nuestro análisis descriptivo.

En tercer lugar, tras la realización del análisis descriptivo, exportaremos nuevamente nuestro fichero de datos al software Eviews12 para estimar nuestro modelo logit, redactando con carácter previo una serie de fundamentos teóricos para comprender en qué consiste dicha estimación.

En cuarto lugar, tras la interpretación de la estimación de nuestro modelo logístico, vamos a cuantificar el efecto parcial de nuestras variables independientes sobre nuestra variable dependiente (encontrarse desempleado).

En quinto lugar, calcularemos las probabilidades de estar desempleado en 32 individuos de nuestra muestra con diferentes cualidades y así crearemos una serie de escenarios ficticios.

Por último, elaboraremos una serie de gráficos (curvas de respuesta) para observar las probabilidades de estar desempleado para cada categoría de las distintas variables en función de la edad.

4. ANÁLISIS DESCRIPTIVO

4.1 Obtención y tratamiento de los datos

Para poner en funcionamiento este estudio, debemos realizar un análisis descriptivo a partir de los datos de la EPA obtenidos a través del INE, estos datos se corresponden con el cuarto trimestre del año 2022.

Se ha decidido escoger esta fecha porque se quiere valorar la situación del desempleo en España en la que nos encontramos actualmente.

El objetivo principal de este estudio radica en explicar los determinantes del desempleo a partir de una muestra de 128.192 encuestados, sin embargo, procedemos a suprimir a aquellos individuos que se encuentran inactivos, ya que, aunque se encuentren desempleados, tampoco buscan empleo de forma activa debido a sus características. Por lo tanto, reducimos dicha muestra a 58.936 individuos.

Haciendo referencia a las variables con las que vamos a trabajar, también vamos a eliminar la gran mayoría que presenta la Encuesta de Población Activa, de tal manera que nuestras variables independientes con la que vamos a trabajar son: la edad, el sexo, la nacionalidad, el nivel de formación y el estado civil.

Cabe destacar que la mayoría de nuestras variables independientes a estudiar se tratan de variables cualitativas. Esto supone que nuestro modelo va a estar restringido desde un primer momento, ya que no vamos a tener en consideración variables cuantitativas relevantes como el salario.

Por otra parte, nuestra variable dependiente va a ser la situación laboral que presentan los encuestados, la cuál va a consistir principalmente entre si están ocupados o parados.

4.2 Recodificación de las variables

La variable *DESEMPLEO* [*AOI*]⁸ es nuestra principal variable, ya que hace referencia a la situación laboral de los encuestados. Vamos a trabajar con ella de la siguiente manera:

$$AOI \begin{cases} 1 & \text{si se trata de individuos desempleados}^9 \\ 0 & \text{si se trata de individuos ocupados}^{10} \end{cases}$$

⁸ Hemos decidido suprimir a todos aquellos individuos que se encuentran inactivos para trabajar de manera más efectiva con nuestra variable dependiente.

⁹ Son las personas de 16 o más años que durante la semana de referencia (la anterior a aquella en que se realiza la encuesta) han estado sin trabajo, disponibles para trabajar y buscando activamente empleo.

¹⁰ Son las personas de 16 o más años que durante la semana de referencia han estado trabajando durante al menos una hora a cambio de una retribución en dinero o especie.

La variable *EDAD* es nuestra única variable independiente cuantitativa y representa los rangos de edad de 5 en 5 años.

El primer tramo comprende entre 16 y 19 años de edad, el segundo tramo entre 20 y 24 años de edad y así sucesivamente hasta llegar a los 65 años, que implica que los individuos tienen 65 años o más.

La variable *SEXO* representa el género de los encuestados:

SEXO { 1 si se trata de hombres
0 si se trata de mujeres

La variable *ECIV* representa la situación civil de los individuos encuestados:

ECIV { 1 si se trata de individuos solteros¹¹
0 si se trata de individuos casados

La variable *NAC* hace referencia a la nacionalidad de los encuestados:

NAC { 1 si se trata de individuos con nacionalidad española¹²
0 si se trata de individuos con nacionalidad extranjera

La variable *NFORMAC* hace referencia al nivel de formación:

NFORMAC { 1 si se trata de individuos sin estudios
2 si se trata de individuos con estudios en educación primaria
3 si se trata de individuos con estudios en educación secundaria
4 si se trata de individuos con estudios en bachiller
5 si se trata de individuos con estudios superiores

¹¹ En la Encuesta de Población Activa existe una distinción más completa, se distingue entre viudos y divorciados, sin embargo, hemos decidido agrupar estos individuos dentro de la calificación de soltero ya que el porcentaje de desempleo prácticamente no se ve alterado.

¹² Se incluyen aquellos individuos que presentan la doble nacionalidad.

4.3 Análisis de las variables

Al finalizar la recodificación de los datos, procedemos a hacer un análisis descriptivo de cada una de las variables anteriormente definidas. Para ello, vamos a tomar como referencia comparativa la tasa de desempleo general que nos arroja el INE en el contexto temporal que deseamos estudiar, es decir, en el cuarto trimestre del año 2022.

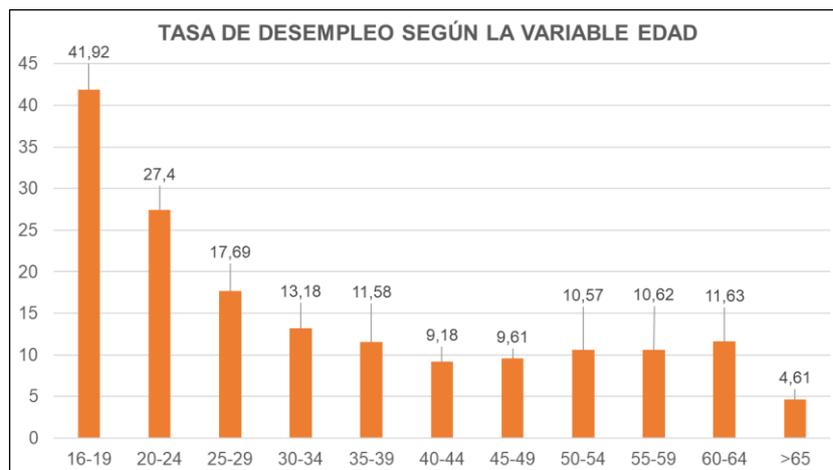
Este desempleo general se encuentra en un 12,2%, este valor se ha reducido en torno a 1 punto absoluto con respecto al cuarto trimestre del año 2021, esta reducción es buena señal ya que venimos de una recesión causada por la crisis sanitaria del año 2020.

Sin embargo, cabe destacar que es muy posible que este 2023 se vea incrementado debido a la actual guerra que se está disputando entre Ucrania y Rusia.

4.3.1 Tasa de desempleo por edades

La primera variable que vamos a analizar es la edad. Tal y como hemos visto en la recodificación de datos, los rangos de edad se agrupan de 5 en 5 años.

Gráfico 4.1 Tasa de desempleo según la variable edad



Fuente: elaboración propia a través de datos consultados en el INE

Podemos apreciar una diferencia notoria entre aquellos individuos en su etapa más joven, es decir, aquellos entre 16 y 29 años frente a aquellos de edad más adulta, mayores de 30 años.

En esta diferencia se puede destacar la tendencia decreciente que sigue el desempleo a medida que se van incrementando los años de edad.

Una de las razones por las que esto puede suceder es por la preferencia que tienen las empresas hacia la experiencia profesional de sus empleados, por lo tanto, es lógico pensar que se contrate a más porcentaje de población dentro del rango de edad que ocupan los adultos.

Por otra parte, es preocupante observar que el desempleo entre los jóvenes es significativamente mayor que el promedio de desempleo general del 12,2%, llegando a valores como 41,92% o 27,40% para aquellos individuos más novatos dentro del mercado laboral.

En cambio, en la edad adulta el desempleo tiende a converger hacia el promedio, mientras que, en la vejez, el desempleo disminuye hasta cuatro veces por debajo del promedio, lo cual es comprensible dado que la mayoría de las personas a esa edad ya se encuentran en su etapa de jubilación.

4.3.2 Tasa de desempleo según el sexo

En la Encuesta de Población Activa, una de las variables a tener en cuenta es el género, el cual se distingue únicamente entre hombres y mujeres. Es importante destacar que esta distinción de género no tiene en cuenta otras identidades de género.

Tabla 4.1 Resumen estadístico de la variable sexo

	Género masculino	Género femenino
Recuento	30.339	28.597
Promedio	10,56%	13,93%

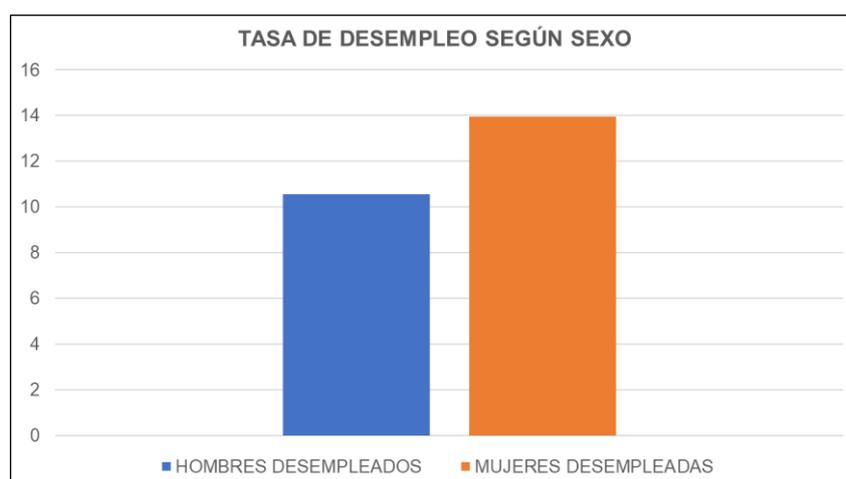
Fuente: elaboración propia a través del software Statgraphics y los datos del INE

Tal y como observamos en la Tabla 4.1 existe una diferencia significativa a la hora de encontrar empleo dependiendo de si eres hombre o mujer. Esta diferencia se encuentra en torno a 3,5 puntos porcentuales en favor del sexo masculino.

Cabe destacar que históricamente las mujeres han enfrentado múltiples barreras y discriminaciones para encontrar empleo dentro del mercado laboral, lo que ha derivado en elevadísimas tasas de desempleo y diferencias enormes entre estos dos tipos de género a lo largo del tiempo.

También es importante volver a mencionar la restricción de este análisis descriptivo, ya que las razones por la que esta diferencia de género sucede, se deben a múltiples factores.

Gráfico 4.2 Tasa de desempleo según la variable sexo



Fuente: elaboración propia a partir de los datos consultados en el INE

A partir del Gráfico 4.2, podemos visualizar claramente esa brecha significativa en el desempleo entre el género masculino y femenino. Si bien es cierto que esta diferencia se ha venido disminuyendo con el paso de los años, aún hay que fomentar políticas que promuevan la igualdad de oportunidades sin distinciones de género.

Tal y como hemos mencionado anteriormente, este estudio no se centra en dar conclusiones para tratar de disminuir esas tasas de desempleo, ya que, para ello, habría que analizar un análisis mucho más detallado y específico que considerase, entre otros muchos factores, la desigualdad salarial.

4.3.3 Tasa de desempleo según la nacionalidad

Vamos a comprobar si la variable Nacionalidad es significativa con respecto a estar desempleado o no estarlo.

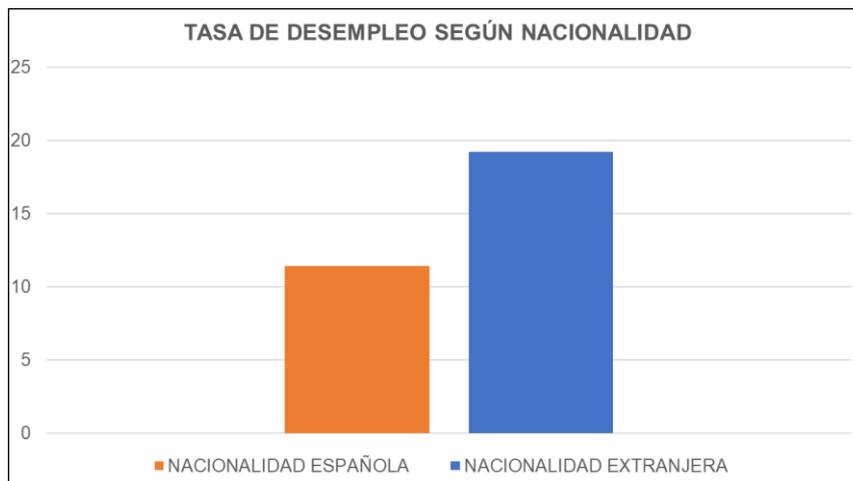
Tabla 4.2 Resumen estadístico de la variable nacionalidad

	Nacionalidad española	Nacionalidad extranjera
Recuento	54.281	4.655
Promedio	11,60%	19,20%

Fuente: elaboración propia a través del software Statgraphics y los datos del INE

La tasa de desempleo entre individuos que poseen la nacionalidad española y los individuos extranjeros refleja una disparidad muy significativa a la hora de encontrar empleo. Esta diferencia en torno a 8 puntos porcentuales puede deberse a las numerosas barreras a las que se pueden enfrentar aquellos individuos que poseen la nacionalidad extranjera, algunos ejemplos pueden ser el acercamiento tanto al idioma y la cultura como al entorno empresarial y legislativo.

Gráfico 4.3 Tasa de desempleo según la variable nacionalidad



Fuente: elaboración propia a partir de los datos consultados en el INE

4.3.4 Tasa de desempleo según el nivel de formación

El nivel de formación que presentan los encuestados es una de las variables seleccionadas para trabajar en nuestro estudio, vamos a comprobar si es significativa para encontrar empleo.

Tabla 4.3 Resumen estadístico de la variable nivel de formación

	Sin estudios	Primaria	ESO	Bachiller	Superiores
Recuento	148	2.540	16.317	14.052	25.878
Promedio	37,16%	27,04%	16,74%	12,9%	7,35%

Fuente: elaboración propia a través del software Statgraphics y los datos del INE

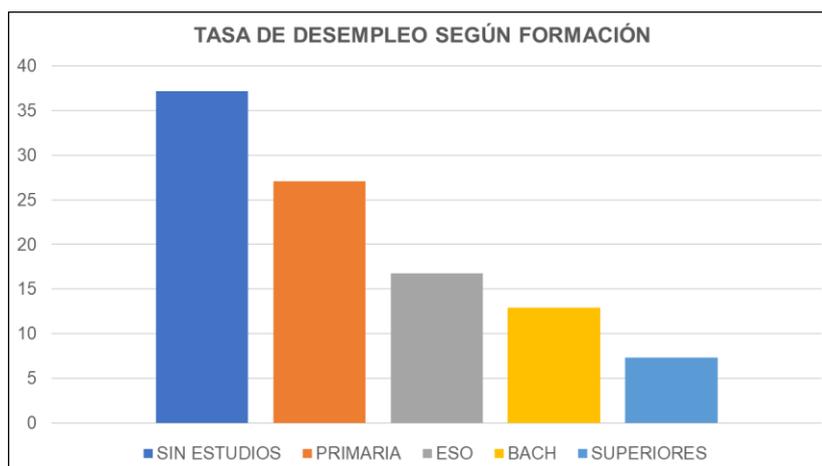
A través de este resumen estadístico podemos observar una clara relación entre el nivel de formación y el desempleo, de tal manera que cuanto mayor sea la formación de un individuo menor será su tasa de desempleo.

Cabe destacar que la tasa de desempleo más baja la presentan aquellos individuos con estudios superiores, esto se debe a la preparación y acercamiento al mundo laboral que suponen. Además, es alentador que un gran porcentaje de los encuestados de la muestra poseen este tipo de estudios.

También es importante destacar la importancia de adquirir un buen nivel de estudios para conseguir empleo en el mercado laboral, ya que, el promedio de los cuatro niveles educativos que se encuentran por debajo del nivel correspondiente a alcanzar estudios superiores¹³ representados en la Tabla 4.3, están por encima del promedio general de desempleo (12,2%) anteriormente mencionado.

¹³ Según la CNED14-P los estudios superiores abarcan enseñanzas de formación profesional, estudios universitarios y enseñanzas de doctorado.

Gráfico 4.4 Tasa de desempleo según la variable nivel de formación



Fuente: elaboración propia a partir de los datos consultados en el INE

Si nos detenemos en el Gráfico 4.4 podemos observar de manera más clara esa relación inversa existente entre el desempleo y el nivel educativo. Siguiendo con lo anteriormente mencionado es importante continuar invirtiendo en educación y formación ya que es un servicio público indispensable y que está en un cambio constante debido a las distintas actuaciones políticas.

A continuación, vamos a añadir la variable nacionalidad para comprobar si existen variaciones en el nivel de formación. En este caso hemos decidido evaluar únicamente a aquellos individuos que poseen, al menos, estudios en bachillerato.

Tabla 4.4 Resumen estadístico de las variables nacionalidad y nivel de formación

	Españoles >= Bach ¹⁴	Extranjeros >= Bach ¹⁵
Recuento	35.955	2.530
Promedio	8,54%	16,60%

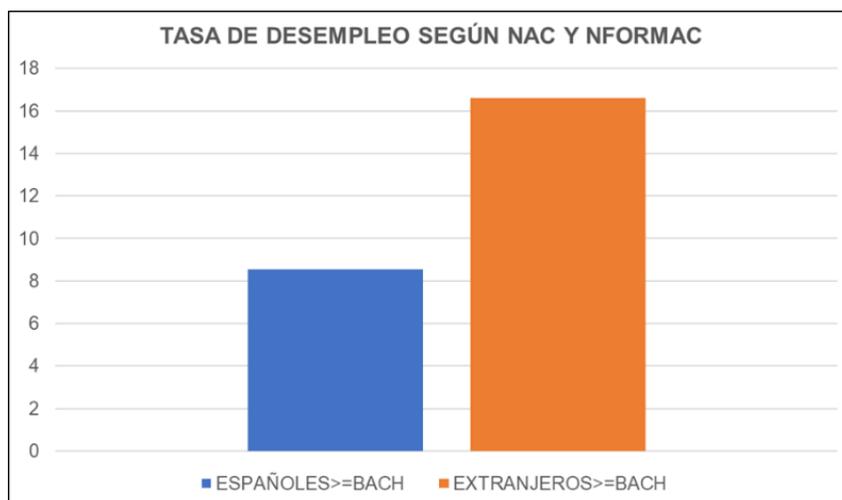
Fuente: elaboración propia a través del software Statgraphics y los datos del INE

¹⁴ Estudiantes que poseen nacionalidad española y han completado al menos la educación en bachillerato

¹⁵ Estudiantes que poseen nacionalidad extranjera y han completado al menos la educación en bachillerato

Podemos observar que la diferencia en la tasa de desempleo se sitúa en torno a 8 puntos porcentuales, de manera similar a lo que observamos al analizar únicamente la variable nacionalidad (ver Tabla 4.2). Sin embargo, podemos ver una clara disminución de 3 puntos porcentuales en cada tasa de paro debido a la introducción del nivel de estudios en bachiller y superiores.

Gráfico 4.5 Tasa de desempleo según las variables nacionalidad y nivel de formación



Fuente: elaboración propia a partir de los datos consultados en el INE

4.3.5 Tasa de desempleo según el estado civil

A lo largo de nuestro análisis descriptivo hemos comprobado que las variables seleccionadas para nuestro estudio están, de una manera o de otra, relacionadas con la tasa de desempleo. A continuación, vamos a comprobar si nuestra última variable, el Estado Civil que presentan los encuestados, incide significativamente en la posibilidad de encontrar trabajo en el mercado laboral.

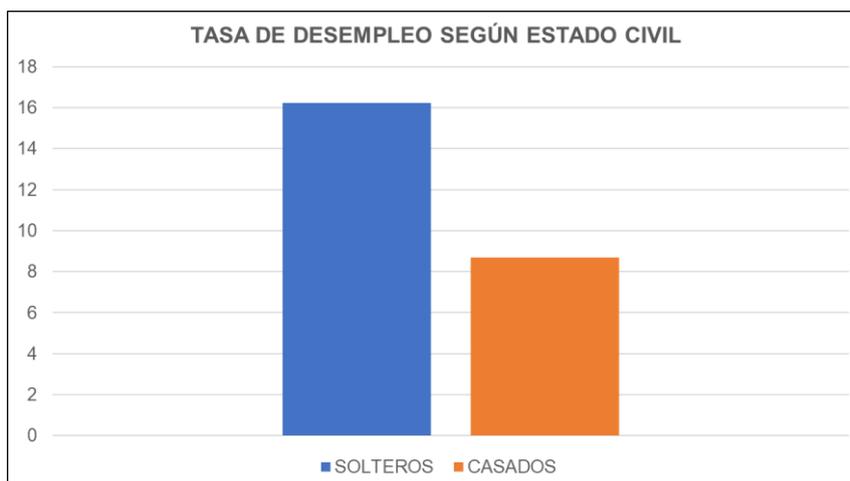
Tabla 4.5 Resumen estadístico de la variable estado civil

	Solteros	Casados
Recuento	27.449	31.468
Promedio	16,22%	8,69%

Fuente: elaboración propia a través del software Statgraphics y los datos del INE

Los resultados que nos arroja este resumen estadístico nos revelan que el Estado Civil incide a la hora de encontrar empleo, de tal forma que para aquellos individuos que presentan la situación de estar casados tienen más posibilidades de encontrar empleo frente a los solteros.

Gráfico 4.6 Tasa de desempleo según la variable estado civil



Fuente: elaboración propia a partir de los datos consultados en el INE

Podemos observar en este gráfico como la diferencia en las tasas de desempleo es abrumadora, siendo la tasa en individuos solteros prácticamente el doble que la tasa en individuos casados.

Una de las razones por la que esta situación sucede puede deberse a que las personas casadas, a menudo tienen responsabilidades familiares, como, por ejemplo, el mantenimiento de algún hijo; esto supone que prioricen la estabilidad y la seguridad en sus empleos.

Además, es razonable pensar que esta diferencia está directamente relacionada con la edad, ya que aquellos individuos más jóvenes presentan con mayor frecuencia la situación civil de soltero, y como hemos visto anteriormente, a menor edad, mayor tasa de desempleo.

Por otra parte, en el caso de los solteros, aunque presentan una mayor flexibilidad laboral, la cual es una de las características más relevantes para encontrar empleo, también es posible que acepten trabajos con peores condiciones laborales y tengan la necesidad de cambiar de empleo con más frecuencia. Esto podría afectar negativamente a la hora de encontrar empleo.

A continuación, vamos a considerar otras variables dentro de nuestra ecuación, esto puede cambiar significativamente la situación. En este caso, se añadirá la variable sexo al estudio. Los resultados obtenidos se pueden examinar en el siguiente resumen estadístico.

Tabla 4.6 Resumen estadístico de las variables estado civil y sexo

	Hombres casados	Mujeres casadas	Hombres solteros	Mujeres solteras
Recuento	16.439	15.048	13.900	13.549
Promedio	5,98%	11,64%	15,97%	16,48%

Fuente: elaboración propia a través del software Statgraphics y los datos del INE

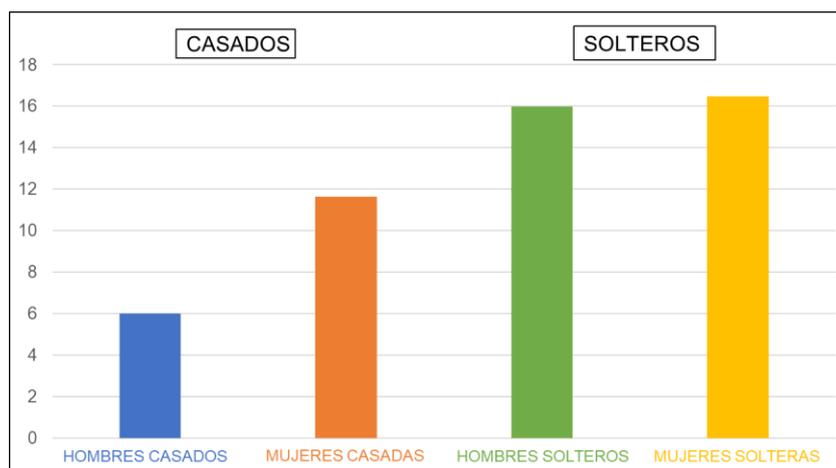
Las tasas de desempleo por sexo son muy diferentes dependiendo de la situación civil en la que se encuentren los individuos, de tal forma que para los casados existe una diferencia en torno a 5,5 puntos porcentuales mientras que para los solteros se encuentra en 0,5 puntos porcentuales.

Podemos observar como la situación cambia con respecto a la Tabla 4.5, ya que para aquellos individuos que se encuentran casados y poseen el género masculino tienen una tasa de desempleo menor al promedio general entre individuos que se encuentran casados, mientras que en el género femenino se encuentran por encima de ese promedio del 8,69%.

Además, hemos querido comprobar si existen diferencias de género entre aquellos encuestados solteros, podemos observar que ambas tasas de desempleo son superiores frente a estar casado (ver Tabla 4.5).

Sin embargo, es interesante observar cómo no existe prácticamente diferencia de género entre los individuos solteros, encontrándose la tasa de desempleo para mujeres ligeramente por encima con respecto a los hombres.

Gráfico 4.7 Tasa de desempleo según las variables estado civil y sexo



Fuente: elaboración propia a partir de los datos consultados en el INE

Esta situación la podemos ver en el Gráfico 4.7. Atendiendo al ámbito de personas que se encuentran casadas, podríamos decir que una de las razones por las que esta diferencia sucede, puede deberse a que las empresas buscan personal con total dedicación a la entidad. Esto se ve mermado en muchas ocasiones por los embarazos en las mujeres, ya que su rendimiento puede verse afectado tanto en el periodo de gestación como después del parto.

Con respecto a esta última afirmación, cabe destacar que la nueva modificación del Real Decreto¹⁶ publicada en el Boletín Oficial del Estado con fecha 01/03/2023 por parte del gobierno puede afectar en la reducción de esta significativa diferencia.

5. MODELO LOGIT

Antes de comenzar a trabajar con nuestro modelo logit, vamos a introducir una serie de conocimientos teóricos necesarios para entender la estructura del modelo.

En primer lugar, vamos a proceder con la construcción de nuestro modelo econométrico.

¹⁶ Real Decreto 295/2009, de 6 de marzo, por el que se regulan las prestaciones económicas del sistema de la Seguridad Social por maternidad, paternidad, riesgo durante el embarazo y riesgo durante la lactancia natural.

Nuestra variable dependiente Y_i es una variable dicotómica, es decir, una Bernoulli que toma dos valores posibles:

$$Y_i \begin{cases} 1 & \text{si el individuo está desempleado; } P_i = P(Y = 1|x) \\ 0 & \text{si el individuo está ocupado; } 1 - P_i = P(Y = 0|x) \end{cases}$$

Al tratarse de una Bernoulli, su esperanza matemática se calcula como la suma ponderada de sus posibles valores, donde cada valor se multiplica por su respectiva probabilidad:

$$E(y_i|x) = 1 * P_i + 0 * (1 - P_i) = P(Y = 1|x)$$

En nuestro modelo, vamos a utilizar un conjunto de regresores X_j , que son las variables independientes que hemos mencionado en nuestro análisis descriptivo previo, vamos a suponer una relación lineal entre la probabilidad de P_i y las variables independientes:

$$P(y_i = 1|x) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_{ki}$$

Por lo tanto, vamos a construir un modelo de regresión lineal múltiple.

$$Y_i = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_{ki} + u_i$$

Por consiguiente, nos queda un MPL¹⁷

$$Y_i = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_{ki} + u_i$$

Sin embargo, al estimar nuestro modelo de probabilidad lineal, nos enfrentamos a un problema importante que afecta a nuestro estudio. Resulta que la probabilidad de estar desempleado puede generar valores negativos y superiores a la unidad, lo cual no es coherente ni interpretable en nuestro caso.

¹⁷ Modelo de probabilidad lineal, consiste en un modelo de regresión lineal múltiple donde los coeficientes beta miden el cambio en la probabilidad de éxito cuando uno de los regresores aumenta una unidad y el resto permanece constante.

Para resolver esto, vamos a utilizar una regresión logit. Se trata de un modelo de regresión no lineal diseñado de manera específica para trabajar con variables dependientes binarias, utilizando una función de distribución de probabilidad acumulada, más concretamente, la función de distribución logística estándar. Por lo tanto, un modelo de regresión logística se representa a partir de la siguiente ecuación.

$$Y_i = F(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_{ki}) + u_i$$

Donde $0 \leq F \leq 1$

Y si aplicamos la función logística estándar tenemos el modelo logit:

$$Y_i = \frac{1}{1 + e^{-(z)}} + u_i$$

Donde z es $\beta_0 + \beta_1 X_1 + \dots + \beta_k X_{ki}$

Por lo tanto, a la hora de estimar el modelo nos queda la siguiente la expresión:

$$\hat{P}_i = p(\widehat{y_i = 1|x}) = \frac{1}{1 + e^{-(\hat{z})}}$$

Donde \hat{z} es $\hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_k X_{ki}$

La interpretación de los coeficientes beta, al tratarse de un modelo no lineal, es diferente con respecto de un modelo de regresión lineal. Si Δx_i en una unidad y el resto permanece constante obtenemos lo siguiente:

$$\Delta \hat{P}_i = F(\hat{\beta}_0 + \hat{\beta}_1(\hat{X}_1 + 1) + \dots + \hat{\beta}_k \hat{X}_{ki}) - F(\hat{\beta}_0 + \hat{\beta}_1 \hat{X}_1 + \dots + \hat{\beta}_k \hat{X}_{ki}) \neq \hat{\beta}_i$$

Si calculamos el efecto marginal:

$$\frac{d\hat{P}_i}{dx_{ij}} = \frac{\partial \hat{P}_i}{\partial z} * \frac{\partial z}{\partial x_{ij}} = \beta_j * f(-x_i' \beta)$$

Por lo tanto, β_j depende del valor de todos los regresores del modelo.

La expresión del efecto marginal la podemos expresar en función de la probabilidad estimada de la siguiente manera:

$$\frac{\partial \hat{P}_i}{\partial x_{ij}} = \hat{P}_i * (1 - \hat{P}_i) * \hat{B}_j$$

Antes de comenzar con la estimación del modelo vamos a introducir un concepto muy utilizado en este tipo de regresiones para interpretar los coeficientes beta, los *odds ratio* (OR) o cocientes entre *odds*.

Definimos un *odd* como:

$$Odd = \frac{p(y = 1)}{p(y = 0)} = \frac{p}{1 - p}$$

$$Donde p = \frac{e^{\beta_0 + \beta_i X_i + \dots + \beta_k X_{ki}}}{1 + e^{\beta_0 + \beta_i X_i + \dots + \beta_k X_{ki}}}$$

Por lo tanto, si operamos obtenemos:

$$Odd = e^{\beta_0 + \beta_i X_i + \dots + \beta_k X_{ki}}$$

Ahora bien, los *odds ratio* miden el cambio en el *odd* cuando un regresor se incrementa en una unidad y el resto permanece constante:

$$Odds\ ratio = \frac{e^{\beta_0 + \beta_i(X_i+1) + \dots + \beta_k X_{ki}}}{e^{\beta_0 + \beta_i X_i + \dots + \beta_k X_{ki}}} = e^{\beta_i}$$

Los *odds ratio* nos van a arrojar unos resultados y dependiendo de qué valor tomen nos encontraremos frente a distintos casos:

Si el OR es mayor que 1, quiere decir que un incremento de una unidad (manteniendo el resto constantes) en una variable explicativa X_j implica un aumento en los *odds* de la variable dependiente. Mientras que cuando el OR es menor que 1, significa que un incremento de una unidad (*ceteris paribus*) en un regresor X_j implica una disminución en los *odds* de la variable dependiente.

Sin embargo, la interpretación para las variables categóricas es distinta, de tal manera que, en vez de medir un incremento, se mide el cambio de categoría entre las variables.

Para OR mayores a 1, el cambio de 0 a 1 en una variable explicativa (*ceteris paribus*) implica un aumento en los *odds* de la variable dependiente y para OR menores a 1, el cambio de 0 a 1 en una variable explicativa (*ceteris paribus*) implica una disminución en los *odds* de la variable dependiente.

5.1 Análisis del modelo

Para realizar el análisis logit es necesario redefinir la variable Nivel de Formación. En nuestro análisis descriptivo previo tomaba valores entre 1 y 5 dependiendo del nivel de formación adquirido por los encuestados. Por otra parte, también deberemos modificar la variable EDAD, de tal manera que cogeremos el punto medio del intervalo de cada uno de los rangos de edad.

La modificación de la variable NFORMAC va a consistir en introducir cuatro variables ficticias¹⁸, de tal manera que, cada una de las variables toman el valor 1 si presentan ese nivel de estudios y 0 si carecen del mismo o si poseen uno superior.

Estas variables serán, PRIM (estudios en educación primaria) ESO (estudios en educación secundaria), BACH (bachillerato), SUP (estudios superiores); la variable correspondiente a no tener estudios la excluirémos del modelo y la utilizaremos como categoría base.

De esta forma, planteamos nuestro modelo logit:

$$AOI_i = F(\beta_0 + \beta_1 EDAD_i + \beta_2 SEXO_i + \beta_3 ECIV_i + \beta_4 NAC_i + \beta_5 PRIM_i + \beta_6 ESO_i + \beta_7 BACH_i + \beta_8 SUP_i) + u_i$$

Donde F es la función de distribución logística estándar.

A continuación, vamos a estimar nuestro modelo logístico, tal y como hemos mencionado anteriormente, al no ser un modelo de regresión lineal no podemos estimarlo por mínimos cuadrados ordinarios (MCO); por lo tanto, vamos a utilizar el método de estimación de máxima verosimilitud¹⁹(EMV).

¹⁸ Introducimos las variables ficticias según la regla general, esto implica que, si tenemos 5 niveles educativos en el nivel de formación, debemos introducir únicamente 4 para no incurrir en la trampa de las variables ficticias, de esta manera evitamos posibles errores en la estimación de nuestro modelo.

¹⁹ Estimadores consistentes y asintóticamente eficientes de los parámetros de nuestro modelo.

Tabla 5.1 Estimación del modelo logit

Dependent Variable: AOI				
Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)				
Date: 06/02/23 Time: 18:48				
Sample: 1 58936				
Included observations: 58935				
Convergence achieved after 6 iterations				
Coefficient covariance computed using observed Hessian				
Variable	Coefficient	Std. Error	z-Statistic	Prob.
EDAD	-0.021733	0.001171	-18.55920	0.0000
SEXO	-0.479256	0.026404	-18.15113	0.0000
ECIV	0.561199	0.028962	19.37716	0.0000
NAC	-0.239667	0.042460	-5.644555	0.0000
PRIM	-0.499350	0.180647	-2.764229	0.0057
ESO	-1.129263	0.177065	-6.377686	0.0000
BACH	-1.575265	0.177727	-8.863397	0.0000
SUP	-2.213389	0.178083	-12.42896	0.0000
C	0.709963	0.183974	3.859037	0.0001
McFadden R-squared	0.068657	Mean dependent var	0.122016	
S.D. dependent var	0.327307	S.E. of regression	0.318261	
Akaike info criterion	0.691217	Sum squared resid	5968.611	
Schwarz criterion	0.692589	Log likelihood	-20359.42	
Hannan-Quinn criter.	0.691643	Deviance	40718.85	
Restr. deviance	43720.59	Restr. log likelihood	-21860.30	
LR statistic	3001.745	Avg. log likelihood	-0.345456	
Prob(LR statistic)	0.000000			

Fuente: elaboración propia a través del programa Eviews

Tal y como hemos mencionado anteriormente en los fundamentos teóricos del modelo logit, los coeficientes en un modelo no lineal se interpretan de manera distinta, de tal forma que vamos a fijarnos únicamente en el signo. Si toman valores negativos, implica que, a medida que aumente el coeficiente, la probabilidad de éxito se verá reducida, es decir, disminuirá la probabilidad de estar desempleado.

En el caso de la variable EDAD a medida que se incremente la edad, la probabilidad de estar desempleado se verá reducida. Para las variables categóricas la interpretación cambia ligeramente, de tal manera que, aquellas que presentan signo negativo en sus coeficientes, cuando la variable cambia del valor 0 al 1, la probabilidad de estar desempleado disminuirá. Por ejemplo, si nos centramos en la variable NAC, pasar de tener la nacionalidad extranjera (0) a poseer la nacionalidad española (1) implica una reducción en la probabilidad de desempleo. Por otro lado, para la variable ECIV, cuando una persona cambia de estar casada (0) a estar soltera (1) implica que la probabilidad de desempleo se incrementará, esto se debe al signo positivo del coeficiente.

Si nos detenemos en la columna "Prob" (ver

Tabla 5.1) podemos observar cómo los p-valor asociados a las variables son $<0,05$. Esto implica que todas las variables son significativas a nivel individual, por lo tanto, no será necesario eliminar ninguna de nuestras variables del modelo. Además, el p-valor correspondiente a la significación conjunta situado en la fila “Prob(LR statistic)” es cero, por lo que también existe significación conjunta entre las variables.

A continuación, vamos a verificar el poder explicativo de nuestro modelo. El “McFadden R-squared” toma valores próximos a cero, esto es indicativo de que nuestro modelo no tiene buen poder explicativo, sin embargo, esta medida de ajuste funciona de manera distinta al coeficiente de determinación R^2 utilizado en regresiones lineales; de tal manera que el “McFadden” funciona principalmente en los modelos logit como herramienta comparativa entre distintas especificaciones de modelos y no como indicador del poder explicativo del modelo.

Por lo tanto, vamos a utilizar los anteriormente mencionados *odds ratio* para interpretar la relación existente entre nuestras variables explicativas y la variable dependiente.

Tabla 5.2 Odds ratio

Variables	Coefficientes β	Odds ratio ²⁰
EDAD	-0,021733	0,978501
SEXO	-0,479256	0,619244
ECIV	0,561199	1,752773
NAC	-0,239667	0,78689
PRIM	-0,49935	0,606925
ESO	-1,129263	0,323271
BACH	-1,575265	0,206953
SUP	-2,213389	0,10933
C	0,709963	2,033916

Fuente: elaboración propia

²⁰ Los cálculos de estos *odds ratio* los hemos realizado a través de la función exponencial aplicada a los coeficientes del modelo, es decir, e^{β_i} .

Tal y como vemos en la Tabla 5.2, para los coeficientes con signo negativo los OR son menores a 1 y para el coeficiente correspondiente a la variable estado civil es mayor que 1. Si recordamos lo anteriormente mencionado:

- *Odds ratio* < 1 implican una disminución en los *odds* de la variable dependiente, es decir, la probabilidad de éxito (desempleo) está disminuyendo en comparación con la probabilidad de no ocurrencia (estar empleado)
- *Odds ratio* > 1 implican un incremento en los *odds* de la variable dependiente, es decir, la probabilidad de éxito (desempleo) está aumentando en comparación con la probabilidad de no ocurrencia (estar empleado)

Analicemos la variable EDAD, si se incrementa en una unidad, los *odds* de la variable dependiente disminuyen casi en un 3%, permaneciendo las demás variables constantes.

El caso cambia para las demás variables, ya que, al ser categóricas, no se pueden interpretar con incrementos unitarios, pero si con cambios en las categorías de las variables, por ejemplo, en la variable ECIV, si se cambia del valor 1 (soltero) al valor 0 (casado), los *odds* de la variable dependiente se incrementan en un 75%, permaneciendo las demás variables constantes.

Podemos observar cómo los valores que nos arrojan los *odds ratio* son coherentes, ya que, en el caso del nivel de formación, a mayor nivel de estudios alcanzado, el *odds ratio* se hace más pequeño. Por ejemplo, en los estudios superiores, se reduce en un 90% con respecto a no tener estudios, permaneciendo las demás variables constantes.

5.1.1 Efectos parciales medios de cada variable

Para cuantificar los efectos parciales medios de nuestras variables explicativas de forma aislada, vamos a estimar el efecto marginal utilizando la expresión anteriormente definida:

$$\frac{\partial \hat{P}_i}{\partial x_{ij}} = \hat{P}_i * (1 - \hat{P}_i) * \hat{B}_j$$

Lo que resta, sería calcular la probabilidad estimada (\hat{P}_i) de cada individuo de la muestra y la obtendremos sustituyendo en:

$$\frac{1}{1 + e^{-(0,709 - 0,021X_1 - 0,479X_2 + 0,561X_3 - 0,239X_4 - 0,499X_5 - 1,129X_6 - 1,575X_7 - 2,213X_8)}}$$

Es importante destacar que este efecto marginal solo lo podemos analizar para variables continuas, es decir, en nuestro caso, únicamente a la variable EDAD. Por lo tanto, una vez realizado el cálculo de las correspondientes probabilidades, sustituimos esos valores en la expresión del efecto marginal y multiplicamos por el coeficiente beta estimado de la variable EDAD, es decir, -0,021.

Por último, calcularemos el promedio de los efectos marginales de cada individuo de la muestra y obtenemos el efecto marginal medio de esta variable. Cuando se incrementa en una unidad los años de edad, la probabilidad de estar desempleado se ve disminuida en 0,00216 puntos porcentuales, manteniendo constantes todas las demás variables. Podemos observar que esta disminución es muy reducida, sin embargo, es lógico pensar que la probabilidad de estar desempleado al incrementarse un año de edad no sea muy significativa.

A continuación, vamos a calcular los efectos medios de las demás variables de nuestro estudio, sin embargo, al tratarse de variables cualitativas categóricas no podremos utilizar el efecto marginal anteriormente mencionado, esto es evidente, ya que es imposible medir incrementos unitarios en variables que toman únicamente dos valores. Por lo tanto, vamos a calcular el efecto medio de una variable al pasar de una categoría a otra, manteniendo las demás variables en su media. Para ello vamos a calcular la diferencia en la probabilidad de estar desempleado cuando las variables toman el valor 1 y el valor 0, manteniendo las demás variables en su promedio.

Vamos a ilustrarlo con un ejemplo utilizando la variable SEXO:

$$\begin{aligned} \hat{P}_i(y = 1|x_2 = 1) &= F(\hat{\beta}_0 + \hat{\beta}_1\bar{x}_1 + \hat{\beta}_2 + \hat{\beta}_3\bar{x}_3 + \dots + \hat{\beta}_k\bar{x}_k) - \\ - \hat{P}_i(y = 1|x_2 = 0) &= F(\hat{\beta}_0 + \hat{\beta}_1\bar{x}_1 + \hat{\beta}_3\bar{x}_3 \dots + \hat{\beta}_k\bar{x}_k) \end{aligned}$$

El primer término $[\hat{P}_i(y = 1|x_2 = 1)]$ hace referencia a la probabilidad de estar desempleado siendo un varón con un 0,0842, mientras que el segundo término $[\hat{P}_i(y = 1|x_2 = 0)]$ hace referencia a la probabilidad de estar desempleado siendo una mujer con un 0,129. El efecto medio en este caso sería de 4,5 puntos porcentuales, esto quiere decir que las mujeres son un 4,5% más propensas a quedar desempleadas con respecto a los hombres, manteniendo el resto constantes.

Tabla 5.3 Efectos medios al cambiar de categoría en las variables cualitativas

Variables	$P(y = \widehat{1} x_k = 1)$	$P(y = \widehat{1} x_k = 0)^{21}$	Efecto medio	Efecto NFORM ²²
SEXO	0,084	0,129	0,045	-
ECIV	0,135	0,082	0,053	-
NAC	0,102	0,126	0,025	-
PRIM	0,274	0,384	0,11	-
ESO	0,167	0,384	0,216	0,107
BACH	0,114	0,384	0,27	0,053
SUP	0,063	0,384	0,32	0,05

Fuente: elaboración propia

Haciendo referencia al estado civil y la nacionalidad, la diferencia en el desempleo se encuentra en un 5,3% en favor de los hombres casados y un 2,5% más probable de quedar desempleado para los individuos extranjeros con respecto a los que poseen la nacionalidad española.

²¹ Para las variables referentes al nivel de formación, cuando valen cero, implica que no se tienen estudios.

²² Hace referencia al efecto en el nivel de estudios alcanzado con respecto al nivel de estudios anterior.

En cuanto al nivel de formación, hemos calculado el efecto parcial de cada una de nuestras variables con respecto a la categoría base (sin estudios), de tal manera que, cuanto más nos acercamos a tener estudios superiores, mayor se hace la diferencia con respecto a no tener estudios. Además, hemos querido calcular las diferencias entre poseer un nivel de estudios con respecto al nivel de estudios anterior; los resultados nos ofrecen una diferencia notoria entre aquellos que han cursado la ESO con respecto a los que únicamente han cursado la educación primaria, más concretamente, aquellos que poseen los estudios en secundaria tienen un 10% más de probabilidades de encontrar empleo. Por otro lado, las diferencias entre los siguientes niveles de estudios se reducen a la mitad.

Sin embargo, este análisis sobre los efectos al cambiar de categoría tiene sus deficiencias, una de ellas es el hecho de utilizar el promedio en nuestras variables cualitativas, ya que, tal y como hemos ido mencionando a lo largo de nuestro estudio, toman únicamente el valor 1 ó 0 y no valores medios.

Es por esto que vamos a plantear una serie de escenarios para ver de manera más clara y concisa las diferencias a la hora de encontrar empleo.

5.1.2 Creación de escenarios

Para la elaboración de los distintos escenarios, hemos decidido analizar a 32 individuos con distintas características para ver cómo pronostica la probabilidad de estar desempleado nuestro modelo (ver Anexo 1).

La distribución de los escenarios se ha dividido de cuatro en cuatro individuos, donde la edad se va incrementando a medida que se avanza un bloque y la diferencia de género se encuentra siempre presente. Cabe destacar que el último bloque se compone de únicamente dos individuos con características muy dispares entre sí.

Hemos comenzado por aquellos individuos jóvenes que tienen 18 años, podemos observar como todos ellos tienen elevadas probabilidades de estar desempleados debido a su muy temprana edad.

También es importante mencionar las descomunales diferencias que surgen tanto por el nivel de formación como por la diferencia de género. Vamos a ilustrarlo con un ejemplo concreto, para un joven español soltero de 18 años que ha completado sus estudios en bachillerato, se estima una probabilidad de 19,55 puntos porcentuales, mientras que para una joven española soltera de 18 años con educación secundaria la probabilidad asciende hasta 38,01 puntos porcentuales.

Continuando con el siguiente bloque de individuos jóvenes, las probabilidades de estar parado siguen siendo elevadas, pero se han reducido con respecto los jóvenes anteriores. En este caso, hemos querido comprobar el efecto de la variable nacionalidad, habiendo una diferencia de 4 puntos porcentuales entre los hombres y de 5 entre las mujeres, ambas diferencias en favor de aquellos que poseen la nacionalidad española.

Finalizando con el bloque de los jóvenes, podemos observar como las probabilidades comienzan a converger con la tasa de desempleo general²³, esto se debe a que estos individuos han alcanzado los estudios superiores. Sin embargo, es preocupante observar como las mujeres con características similares siguen teniendo una probabilidad de desempleo pronosticada superior a la tasa de desempleo general. Para el caso de una mujer soltera de 27 años, española y con estudios superiores alcanza un 14,57% de probabilidad estimada, mientras que para el mismo caso variando únicamente la nacionalidad asciende a 17,81 puntos porcentuales. Ambos casos por encima de la tasa general situada en el 12,2%.

Hasta ahora no hemos analizado la incidencia que puede suponer el estado civil en la probabilidad de encontrarse desempleado, esto es debido a que, en la población más joven, abundan los individuos solteros.

Los siguientes doce escenarios hacen referencia a los individuos en su etapa adulta, comprendiendo edades entre 37 y 47 años. En este primer bloque podemos apreciar una diferencia notoria en la probabilidad pronosticada, obteniendo mayores posibilidades de empleo los individuos casados.

²³ La tasa de desempleo general utilizada en nuestra muestra estaba fijada, según el INE en el cuarto trimestre del año 2022, en un 12,2%.

La situación más favorable hasta ahora se corresponde con un hombre casado de 37 años, español y con estudios en bachillerato con una probabilidad de desempleo estimada del 8,4%.

En el siguiente bloque, se ha decidido comparar el nivel de formación al mismo tiempo que la nacionalidad, encontrándose los cuatro individuos casados. Las probabilidades de estar parado se encuentran por debajo de la tasa de desempleo general, además, podemos apreciar como el hecho de alcanzar estudios superiores brindan más oportunidades a la hora de encontrar empleo. Vamos a ilustrarlo con un ejemplo, no existen diferencias significativas en las probabilidades de encontrarse en paro entre un hombre casado español con estudios en bachiller (7,69%) y una mujer casada extranjera con estudios superiores (8,19%).

Dentro de los últimos seis escenarios nos encontramos a los individuos de edad más adulta, podemos apreciar las oportunidades de encontrar empleo más elevadas dentro de nuestro análisis probabilístico para aquellos individuos que han alcanzado estudios superiores; en contraposición, podemos observar como aquellos que no los han alcanzado se encuentran en una situación desfavorable para encontrar empleo.

Por último, hemos querido comparar el escenario más desfavorable con el más favorable, de tal manera que, para una mujer de 18 años soltera, extranjera y sin estudios su probabilidad de estar en desempleo asciende hasta un 70%, mientras que, para un hombre de 65 años casado, español y con estudios superiores su probabilidad de estar en desempleo se encuentra en un 2,5%.

En resumen, podemos concluir que existen varios factores que influyen en la probabilidad de estar desempleado. Uno de los aspectos más relevantes es el nivel de educación alcanzado, de tal manera que, a mayor nivel de estudios, la probabilidad de estar desempleado disminuye.

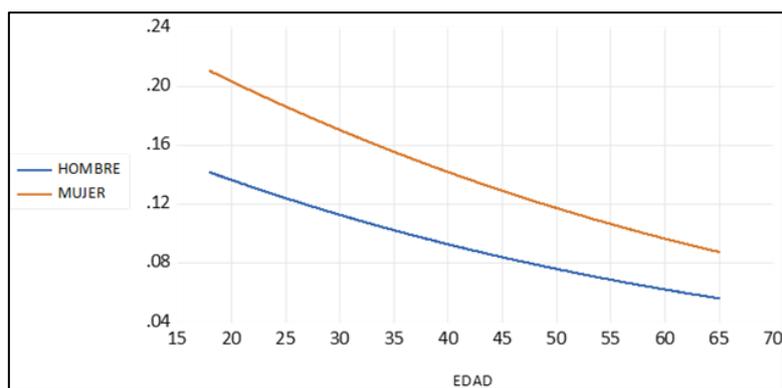
Además, el estado civil que presentan los individuos también desempeña un papel importante, de tal forma que los individuos casados presentan mejores resultados.

La edad es otro factor determinante, ya que, a medida que se van incrementando los años de edad, a priori, se va disminuyendo la probabilidad de encontrarse desempleado, además se puede observar como las diferencias en cuanto al género y nacionalidad son más pronunciadas en edades más jóvenes. Sorprendentemente, la variable nacionalidad, tiene una menor influencia en comparación con las otras variables.

5.1.3 Curvas de respuesta

Por último, vamos a analizar cómo indican nuestras variables en la probabilidad de encontrarse desempleado desde un punto de vista gráfico. Para ello vamos a elaborar una serie de curvas de respuesta, donde representaremos las variables cualitativas en función de la edad.

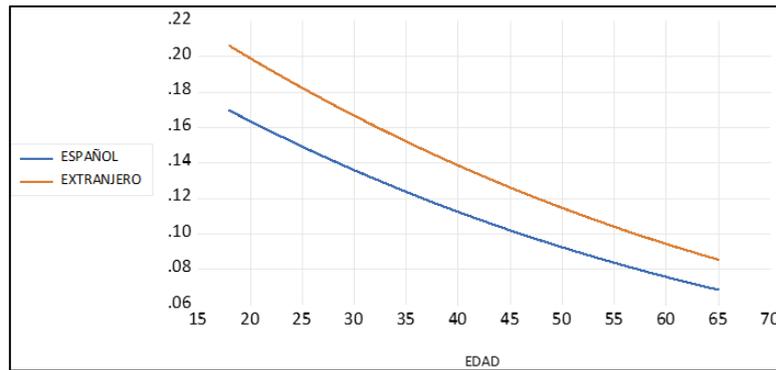
Gráfico 5.1 Curva de respuesta de la variable sexo en función de la variable edad



Fuente: elaboración propia a través del software Eviews12

Este gráfico nos muestra la probabilidad de encontrarse en paro tanto para hombres como para mujeres según la edad alcanzada. Podemos observar que la diferencia de género es significativa a la hora de buscar empleo, siendo más pronunciada en edades más jóvenes y a medida que las personas envejecen, la brecha se reduce, sin embargo, es importante destacar que la probabilidad es siempre superior para el género femenino.

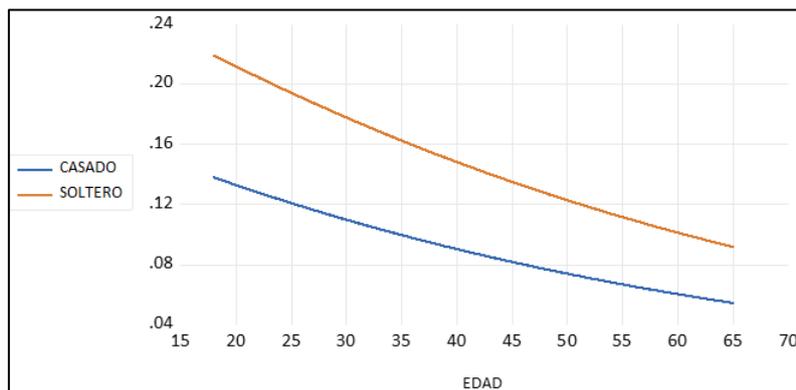
Gráfico 5.2 Curva de respuesta de la variable nacionalidad en función de la variable edad



Fuente: elaboración propia a partir del software Eviews12

Tal y como observamos en el Gráfico 5.2 los individuos que poseen la nacionalidad española tienen más posibilidades a la hora de encontrar empleo con respecto a los extranjeros independientemente de la edad alcanzada. Cabe destacar que la diferencia en la variable nacionalidad es considerablemente menor en comparación con la brecha de género representada en el Gráfico 5.1.

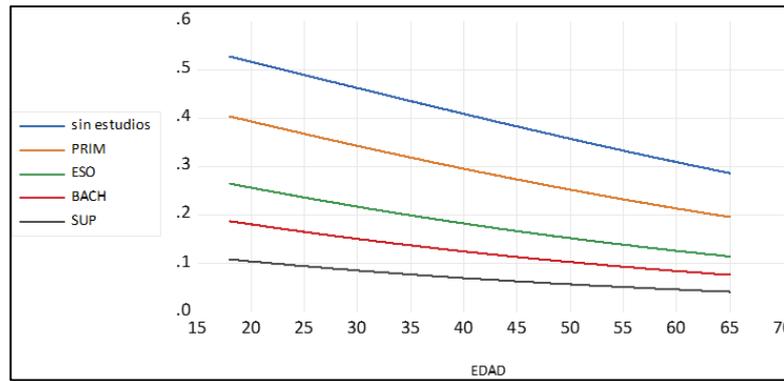
Gráfico 5.3 Curva de respuesta de la variable estado civil en función de la variable edad



Fuente: elaboración propia a partir del software Eviews12

A partir de este gráfico, podemos observar que al igual que en los dos gráficos anteriores, a medida que se van incrementando los años de edad, la probabilidad de encontrarse en paro va disminuyendo. La diferencia en la probabilidad de estar desempleado entre los individuos casados y solteros es sorprendente, siendo en todo momento favorable para los casados, aunque cabe destacar que esta diferencia se va reduciendo a medida que aumenta la edad.

Gráfico 5.4 Curva de respuesta según los niveles de formación en función de la variable edad



Fuente: elaboración propia a partir del software Eviews12

En el caso de los niveles de formación, podemos observar esa tendencia decreciente en la probabilidad de desempleo a medida que se van aumentando los años de edad.

Resulta evidente que esta variable incide de manera considerable a la hora de encontrar empleo. Aquellos individuos que alcanzan una buena formación educativa, especialmente a niveles superiores, se encuentran en situaciones más favorables. Las diferencias más significativas se producen en los niveles educativos más bajos, de tal manera que, a medida que se alcanza mayor formación, las diferencias en la probabilidad de desempleo se van reduciendo.

Por ejemplo, para aquellos individuos que no presentan ningún nivel de estudios, en la etapa más joven, la probabilidad de desempleo se encuentra por encima de 50 puntos porcentuales y a medida que aumenta la edad se va reduciendo de manera considerable; en contraposición, para aquellos que han alcanzado estudios superiores, en la etapa más joven la probabilidad se sitúa alrededor del 10%, y a medida que aumenta la edad, esta probabilidad se va reduciendo, aunque en menor medida en comparación con la situación anteriormente mencionada.

En líneas generales, es importante destacar que, a medida que se alcanzan niveles de estudios más altos, la probabilidad de desempleo en los individuos jóvenes se reduce considerablemente.

6. CONCLUSIONES

Tras finalizar nuestro análisis y estimación del modelo logístico para estudiar los determinantes del desempleo, podemos concluir nuestro estudio habiendo obtenido unos resultados significativos y coherentes.

En primer lugar, hablaremos de los resultados que hemos obtenido gracias al análisis descriptivo previo:

- La edad supone un factor relevante en cuanto al desempleo, de tal manera que, tal y como hemos ido viendo a lo largo del trabajo, existe una relación inversa, que consiste en que a medida que se van incrementando los años, la tasa de desempleo disminuye, estabilizándose cuando el individuo alcanza su etapa más adulta. Por ejemplo, La tasa de desempleo para individuos en el rango de edad de 16 a 19 años alcanza un 41,92%, en contraste con el intervalo de 40 a 44 años, donde la tasa de desempleo se sitúa en un 9,18%.
- Se ha observado una brecha significativa en la tasa de desempleo en las variables sexo, nacionalidad y estado civil, obteniendo menores porcentajes de empleabilidad las mujeres, los extranjeros y los solteros. Estas diferencias son de aproximadamente 3,5 puntos porcentuales en favor de los hombres en términos de género, así como una diferencia de aproximadamente 8 puntos porcentuales tanto para la nacionalidad como para el estado civil, favoreciendo a aquellos que poseen la nacionalidad española y a las personas casadas respectivamente.
- El nivel de formación juega un papel determinante en la incidencia del desempleo, de tal forma que, a medida que se incrementa el nivel de estudios alcanzando, la tasa de desempleo disminuye. Es interesante destacar que aquellos individuos que han alcanzado estudios superiores presentan una tasa excepcionalmente baja, situándose en un 7,35%.

En segundo lugar, atendiendo al modelo logit, la explicación teórica previa a la estimación ha sido esencial para comprender el funcionamiento del mismo. En cuanto a los resultados obtenidos a través de la estimación del modelo, podemos sacar las siguientes conclusiones:

- Todas las variables que hemos seleccionado para nuestro estudio han resultado ser significativas tanto a nivel individual como a nivel colectivo, esto implica que las variables con las que estamos trabajando son relevantes y tienen un impacto considerable sobre nuestra variable dependiente (desempleo).
- Los coeficientes beta estimados presentan signos coherentes. El estado civil es la única variable que presenta signo positivo²⁴, esto implica que, cambiar de estar casado a estar soltero produce un aumento en la probabilidad de desempleo. Las demás variables presentan signos negativos²⁵; es el caso del sexo, la nacionalidad y los cuatro niveles de formación (primaria; secundaria; bachillerato; superiores), cuando estas variables cambian del valor 0 al valor 1, implica que la probabilidad de estar en paro se reduce. Para la variable edad, la interpretación del signo negativo cambia, ya que, al tratarse de una variable continua, un incremento en la edad implica una reducción en la probabilidad de desempleo.

Además, se han llevado a cabo una serie de análisis para comprobar, de manera consistente, las variaciones de nuestras variables independientes en la probabilidad de estar en paro.

El cálculo de los efectos parciales medios nos brinda una aproximación de la probabilidad de encontrarse desempleado para cada una de nuestras variables ceteris paribus. En el caso del sexo, la diferencia en la probabilidad de estar en paro entre hombres y mujeres se encuentra en 4,5 puntos porcentuales en favor del género masculino.

El estado civil nos arroja una diferencia aún más grande que el anterior caso, siendo la probabilidad para los individuos solteros del 13,54% y para los individuos casados un 8,2%.

²⁴ El estado civil es igual a 1 cuando el individuo es soltero y 0 cuando está casado

²⁵ El sexo es igual a 1 (hombre) y 0 (mujer); la nacionalidad es igual 1 (español) y 0 (extranjero); los niveles de formación son iguales a 1 (tienen el nivel de formación que les corresponde a cada uno) y 0 (sin estudios).

También se han decidido crear los escenarios ficticios, los cuales, al utilizar varias combinaciones distintas, nos brindan unos resultados más interesantes, ya que nos permiten pronosticar la probabilidad de encontrarse desempleado dadas las características personales de cada individuo. Se ha decidido comparar la situación más favorable contra la más desfavorable, es el caso de un hombre de 65 años casado, español y con estudios superiores, el cual arroja una probabilidad de desempleo estimada del 2,57%, mientras que en el otro caso se ha estudiado a una mujer de 18 años soltera, extranjera y sin estudios, con una probabilidad estimada del 70,68%.

A raíz de los resultados obtenidos en estos escenarios sacamos las siguientes conclusiones: el nivel de educación alcanzado tiene un impacto significativo en el desempleo, mostrando que, a mayor nivel de estudios, menor es la probabilidad de estar desempleado. El estado civil también influye, con mejores resultados para los individuos casados. A medida que se incrementa la edad, la probabilidad de desempleo disminuye. Las diferencias de género y nacionalidad son más notables en edades más jóvenes, pero la variable nacionalidad tiene una influencia menor en comparación con otras variables.

El último análisis que se ha realizado han sido las curvas de respuesta, a través de estos gráficos, pretendemos analizar la probabilidad de estar desempleado para cada una de las categorías de las diferentes variables cualitativas en función de la edad. Los resultados que nos ofrecen estos gráficos se asemejan a los obtenidos en los escenarios, remarcando la incidencia que supone tanto la edad como el nivel de formación.

Tras realizar todos estos análisis, se ha podido comprobar la importancia que suponen las características personales a la hora de encontrar empleo en el mercado laboral, todas y cada una de las características que se han decidido estudiar han sido significativas en nuestro estudio, remarcando la importante incidencia de la edad y el nivel de formación, así como la sorprendente diferencia en el estado civil y las diferencias no tan notorias pero persistentes del sexo y la nacionalidad.

7. REFERENCIAS BIBLIOGRÁFICAS

Boletín Oficial del Estado. (2010, 30 de marzo). *Versión Consolidada del Tratado de Funcionamiento de la Unión Europea*. Diario Oficial de la Unión Europea. Disponible en: <https://www.boe.es/doue/2010/083/Z00047-00199.pdf> [última consulta: 29/05/2023]

Centro de Investigaciones Sociológicas (CIS) (2014). *El paradigma de la flexiguridad en las políticas de empleo españolas: un análisis cualitativo*. (Vol. 284, pp. 80-91). Madrid: Consejo Editorial de la colección Monografías.

Comisión Europea (2022). *El semestre europeo*. Obtenido de “el paquete de otoño”. Disponible en: [El paquete de otoño \(europa.eu\)](https://ec.europa.eu/economy_finance/) [última consulta: 24/05/2023]

Encarnación Cordón Lagares y Félix García Ordaz (2012). *Principales antecedentes en la consideración del trabajo en la literatura económica: teorías sobre el mercado de trabajo y el desempleo*. Revista Crítica de Historia de las Relaciones Laborales y de la Política Social, nºs 4-5 (marzo 2012/noviembre 2012)

European Commission (2022). *AMECO database*. Disponible en: <https://economy-finance.ec.europa.eu/economic-research-and-databases/economic-databases/ameco-database> en [última consulta: 03/05/2023]

Instituto de Estudios Fiscales (IEF). César Pérez López (2012). *Econometría básica. Aplicaciones con EVIEWS, STATA, SAS y SPSS*. (1ª Edición, pp. 253-259; 310-317). IBERGACETA PUBLICACIONES, S.L.

Instituto Nacional de Estadística (INE) (2022). *Encuesta de Población Activa*. Disponible en: [INEbase / Mercado laboral / Actividad, ocupación y paro / Encuesta de población activa / Resultados](https://inecovid.inecovid.es/INEbase/Mercado%20laboral/Actividad,%20ocupaci%C3%B3n%20y%20paro/Encuesta%20de%20poblaci%C3%B3n%20activa/Resultados) [última consulta: 10/04/2023]

Jeffrey M. Wooldridge (2001). *Introducción a la Econometría. Un enfoque moderno*. (1ª Edición, pp. 529-540). International Thomson Editores, S.A.

James H. Stock y Mark W. Watson (2012). *Introducción a la Econometría*. (3ª Edición, pp. 275-300). Pearson Educación, S.A.

8. ANEXOS

Anexo 1 Distintos escenarios

EDAD	SEXO	ECIV	NAC	NFORMAC	\hat{P}_i
18	Hombre	Soltero	Español	BACH	19,55%
18	Mujer	Soltera	Española	BACH	28,19%
18	Hombre	Soltero	Español	ESO	27,5%
18	Mujer	Soltera	Española	ESO	38,01%
22	Hombre	Soltero	Español	BACH	18,22%
22	Mujer	Soltera	Española	BACH	26,46%
22	Hombre	Soltero	Extranjero	BACH	22,07%
22	Mujer	Soltera	Extranjera	BACH	31,38%
27	Hombre	Soltero	Español	SUP	9,55%
27	Mujer	Soltera	Española	SUP	14,57%
27	Hombre	Soltero	Extranjero	SUP	11,83%
27	Mujer	Soltera	Extranjera	SUP	17,81%
37	Hombre	Soltero	Español	BACH	13,85%
37	Mujer	Soltera	Española	BACH	20,62%
37	Hombre	Casado	Español	BACH	8,4%
37	Mujer	Casada	Española	BACH	12,9%

42	Hombre	Casado	Español	BACH	7,69%
42	Mujer	Casada	Española	BACH	11,73%
42	Hombre	Casado	Extranjero	SUP	5,23%
42	Mujer	Casada	Extranjera	SUP	8,19%
47	Hombre	Soltero	Extranjero	ESO	20,40%
47	Mujer	Soltera	Extranjero	ESO	29,32%
47	Mujer	Soltera	Española	BACH	17,29%
47	Hombre	Soltero	Español	BACH	11,46%
57	Hombre	Soltero	Español	SUP	5,21%
57	Hombre	Casado	Español	SUP	3,24%
57	Mujer	Casada	Española	SUP	4,82%
62	Mujer	Soltera	Española	BACH	13,11%
62	Mujer	Soltera	Extranjera	BACH	16,09%
62	Hombre	Casado	Español	BACH	6,34%
18	Mujer	Soltera	Extranjera	SIN ESTUDIOS	70,68%
65	Hombre	Casado	Español	SUP	2,57%

Fuente: elaboración propia