



---

**Universidad de Valladolid**

**Facultad de Ciencias**

**Económicas y Empresariales**

**Trabajo de Fin de Grado**

**Grado en Marketing e Investigación de  
Mercados**

**Un Modelo Logit para los  
Determinantes del Desempleo en  
España**

Presentado por:

**Alejandro Gómez Velasco**

*Valladolid, 20-07-2022*



## **Resumen**

Este trabajo de fin de grado consiste en la elaboración de un modelo logit que explique los determinantes del desempleo en España. En este trabajo nos hemos centrado en las variables sexo, edad, estado civil, nacionalidad y nivel de estudios. Para ello hemos utilizado los datos de empleo que ofrece el INE del primer trimestre de 2016 y los hemos resumido y codificado para obtener una muestra acorde al objetivo a conseguir. Posteriormente, mediante los programas estadísticos *Statgraphics* y *SPSS* se han realizado distintos análisis en dicha muestra para finalmente obtener el modelo logit que buscábamos el cual hemos usado en la parte final del trabajo para calcular la probabilidad de empleo de diferentes individuos ficticios con distintas características. Gracias a este modelo hemos podido comprobar que es más difícil encontrar trabajo siendo mujer o inmigrante, que conforme la edad del individuo aumenta también lo hacen sus posibilidades de conseguir un trabajo, o que si tenemos en cuenta solo a individuos solteros no hay diferencias de tasa de desempleo entre sexos.

## **Abstract**

This end of degree Project is about making a logit model trying to explain the unemployment determinants in Spain. For this Project we have selected the gender, age, marital status, nationality and educational level. We have used the employment data from the *INE* according to the first trimester of 2016. We have resumed and encoded this data in order to obtain a sample according to our objective. Using the statistical programs *Statgraphics* and *SPSS* we have done some analyses to find the logit model we were looking for. We used this model to estimate the unemployment probabilities for different fictitious people and obtained some conclusions like it is more difficult to find a job being a woman or an immigrant, the possibilities of finding a job increase with the age, or that there is no difference between men and women if we only look into single people.

JEL: C35 - Modelos de regresión discreta y de elección cuantitativa

JEL: C01 – Econometría

JEL: E24 - Empleo; desempleo; salarios; distribución intergeneracional del ingreso; capital humano agregado

## Índice

<b>1. INTRODUCCIÓN.....</b>	<b>1</b>
<b>2. ANÁLISIS DESCRIPTIVO DE LA MUESTRA .....</b>	<b>4</b>
<b>2.1 Obtención y codificación de los datos .....</b>	<b>4</b>
<b>2.2 Realización del análisis descriptivo y resultados .....</b>	<b>5</b>
2.2.1 Análisis de la variable “edad” .....	6
2.2.2 Análisis de la variable “estado civil” .....	7
2.2.3 Análisis de la variable “sexo” .....	8
2.2.4 Análisis de la variable “nacionalidad” .....	10
2.2.5 Análisis de la variable “formación” .....	11
<b>3. MODELO LOGIT.....</b>	<b>13</b>
<b>3.1 Fundamentos teóricos .....</b>	<b>13</b>
<b>3.2. Análisis logit. ....</b>	<b>15</b>
3.2.1 Efecto de cada variable .....	18
<b>4. Escenarios .....</b>	<b>19</b>
<b>4.1 Curvas de respuesta .....</b>	<b>24</b>
<b>5. Conclusiones .....</b>	<b>26</b>
<b>6. Bibliografía.....</b>	<b>28</b>

## Índice de tablas y gráficos

Gráfico 1. 1 Tasa de paro histórica en España (1975-2019).....	1
Gráfico 1. 2 Comparación tasas de paro entre España e Italia .....	2
Gráfico 2. 1 Porcentaje de desempleados según la edad .....	6
Gráfico 2. 2 Tasa de desempleo según el estado civil .....	8
Gráfico 2. 3 Tasa de desempleo según el sexo.....	9
Gráfico 2. 4 Tasa de desempleo según el sexo y el estado civil .....	9
Gráfico 2. 5 Tasa de desempleo según la nacionalidad .....	11
Gráfico 2. 6 Tasa de desempleo según el nivel de estudios.....	12
Gráfico 4. 1 Curva de respuesta edad-sexo.....	24
Gráfico 4. 2 Curva de respuesta edad-estado civil .....	24
Gráfico 4. 3 Curva de respuesta edad-nacionalidad .....	25
Gráfico 4. 4 Curva de respuesta edad-nivel de estudios .....	26
Tabla 2. 1 análisis descriptivo de la variable "estado civil" .....	7
Tabla 2. 2 Análisis descriptivo de la variable "sexo" .....	8
Tabla 2. 3 Análisis descriptivo de la variable "nacionalidad" .....	10
Tabla 2. 4 Análisis descriptivo de la variable "formación" .....	11
Tabla 3. 1 Variables en la ecuación modelo logit .....	15
Tabla 3. 2 Porcentaje de aciertos inicial modelo logit.....	16
Tabla 3. 3 Porcentaje de aciertos final modelo logit .....	17
Tabla 3. 4 Efecto de las variables referentes al nivel de estudios.....	18
Tabla 3. 5 Efecto de las variables "sexo" "estado civil" y "nacionalidad" .....	19
Tabla 4. 1 Distintos escenarios .....	21

## 1. INTRODUCCIÓN

El objetivo de este trabajo es analizar mediante un modelo logit los posibles determinantes del empleo en España. España es un país en el que el desempleo siempre se encuentra en el punto de mira y donde reducirlo es más difícil que en los países de su entorno, ya sea por la terciarización del mercado laboral, por las dificultades de movilidad que tienen los ciudadanos u otros factores.

Para este análisis se ha escogido el primer trimestre del año 2016. A la hora de elegir el año se tuvo en cuenta la actual crisis causada por la COVID-19 y la crisis de 2008. Por lo que se ha intentado elegir un punto en el que el empleo se hubiera recuperado de esta primera crisis pero sin llegar a 2020 donde el empleo cayó mucho por la pandemia que aún vivimos. Se eligió al azar un año entre 2015 y 2019 para que el año fuese completamente aleatorio, pero la elección del trimestre sí que fue condicionada, ya que, como se ha mencionado antes, la terciarización y la temporalidad del mercado laboral podrían afectar si eligiésemos los meses de verano, por lo que se optó por elegir este primer trimestre.



Gráfico 1. 1 Tasa de paro histórica en España (1975-2019)

Como podemos ver en este gráfico, que abarca desde el final de la dictadura en 1975 hasta el 2019, la tasa de paro siempre ha tenido unos niveles bastante elevados, siendo considerado un desempleo menor al 10% como una buena situación. Mientras que si nos fijamos en países de la unión europea como Alemania o Francia serían niveles demasiado elevados y se priorizaría mejorar esa situación. Por ejemplo, el porcentaje de desempleo en Alemania en el año 2010 era menor al 8%, mientras que en España rondaba el 20%, o en 2019, el mejor año desde la crisis de 2008, el desempleo en España se redujo por debajo del 15%, pero en Alemania era de tan solo un 5%.

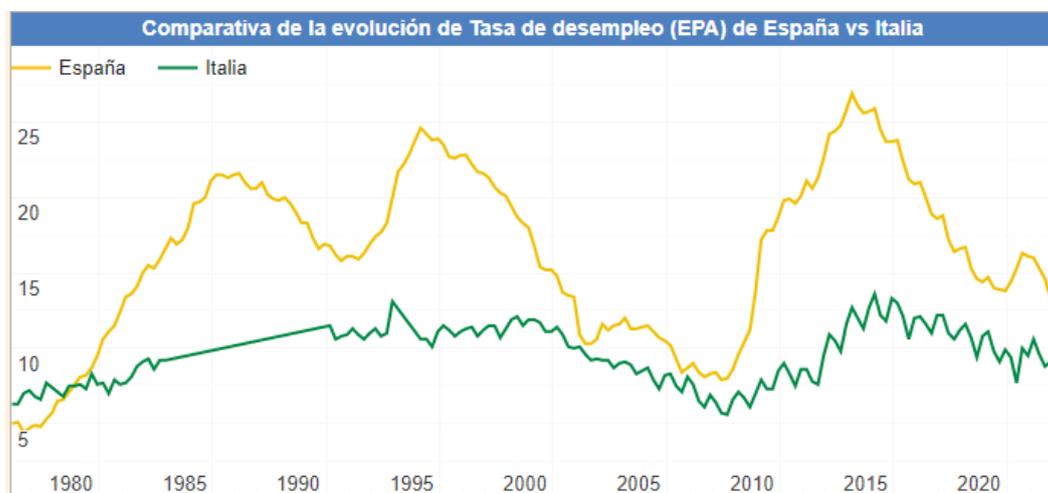


Gráfico 1. 2 Comparación tasas de paro entre España e Italia

En este gráfico podemos apreciar la comparación de las tasas de desempleo española e italiana, países de similares características en cuanto a geolocalización y cultura pero donde se aprecia una clara diferencia entre ambas tasas de desempleo.

Este elevado desempleo puede deberse a varios factores que a continuación explicaremos:

- Elevada tasa de temporalidad: Una gran cantidad de la población activa tiene un contrato temporal, además que la gran mayoría de nuevos contratos que se firman son contratos temporales. Esto se debe a que el costo por despido es mucho menor en este tipo de contratos que en los contratos fijos, por lo que los empresarios optan en su mayoría por esta opción. Esto a su vez provoca en los trabajadores una sensación de desánimo, ya que si consiguen un trabajo con

contrato temporal en un futuro cercano tendrán que buscar empleo de nuevo, el cual a su vez muy posiblemente también será temporal.

- Dificultades a la movilidad: La mayoría de españoles vive en una casa propia a diferencia de otros países europeos o de Estados Unidos donde lo más habitual es vivir de alquiler. Esto hace que los españoles solo busquen trabajo en los alrededores de su zona de residencia, lo cual hace que encontrar trabajo sea más difícil, además de crear grandes desigualdades de desempleo entre comunidades autónomas. Además de este problema con la vivienda también se presentan otras dificultades para buscar trabajo fuera de la comunidad de origen, como pueda ser el idioma propio de algunas comunidades como puedan ser Cataluña o País Vasco. Esto se traduce en una gran desigualdad de la tasa de desempleo, por ejemplo, para el primer trimestre de 2016, la tasa de paro de Andalucía es de casi un 30%, mientras que en el país vasco es de menos de un 13%, una enorme diferencia.

Estos son algunos de los motivos que más saltan a la vista para explicar la situación del desempleo en nuestro país, aunque podría haber muchos otros como el salario, el coste de contratación por empleado o algún motivo cultural.

Teniendo en cuenta esta situación realizaremos un análisis del mercado de trabajo español en busca de conseguir un modelo logit con el que tratar de pronosticar las probabilidades de estar desempleado de un individuo en función de sus características. Las variables elegidas para este estudio han sido la edad, el sexo, el estado civil, la nacionalidad y el nivel de estudios. Debemos tener en cuenta una variable muy importante y que no podemos estudiar como es el salario, por lo que nuestro modelo no será el ideal, aunque con las variables elegidas podemos hacernos una idea bastante aproximada de cómo de fácil será encontrar trabajo para un individuo en función de sus características.

## 2. ANÁLISIS DESCRIPTIVO DE LA MUESTRA

### 2.1 Obtención y codificación de los datos

Lo primero que hemos hecho para la realización de este estudio, ha sido obtener los datos para su posterior codificación. Los datos han sido obtenidos de la página web oficial del Instituto Nacional de Estadística y corresponden al primer trimestre del año 2016. Una vez obtenidos los datos se simplificará la muestra para poder estudiar solo las variables e individuos que nos interesen. Se eliminaron los individuos inactivos de la muestra, quedándonos solo con los ocupados y los parados, que es lo que realmente nos interesa.

Posteriormente se procedió a resumir esos datos, puesto que la información que nos ofrece la tabla descargada de la web del INE es muy abundante y solo nos interesan unas pocas variables concretas. En este caso las variables que hemos elegido han sido la edad, el sexo, el estado civil, la nacionalidad, el nivel de estudios y, por supuesto, la situación laboral.

Una vez hecho este resumen procedemos a la codificación de los datos para poder trabajar cómodamente con ellos. Por ejemplo, en el caso del estado civil, en la tabla que nos ofrece el INE aparecen las siguientes categorías: Soltero, casado, viudo, separado/divorciado. Para simplificar la muestra las variables se han reducido a solo dos, casado y no casado. En la variable “No casado” se incluyen las variables “soltero”, “viudo” y “separado/divorciado”.

A continuación, se explica la codificación realizada en las variables, para que no surjan dudas a la hora de interpretar las tablas de los análisis, así como el nombre que le hemos dado a la variable y su significado.

**AOIB:** Hace referencia a la actividad que desarrolla el encuestado y ha sido codificada de la siguiente manera: 1=desempleado; 0=ocupado

**EDADcont:** hace referencia a la edad mostrada como una variable continua. En el INE nos ofrecen la edad mostrada en rangos de 5 años (variable que veremos a continuación). Lo que hemos hecho ha sido coger para cada dato el valor central del rango, por ejemplo, para el rango que va de los 20 a los 24 años, cogemos el 22. Se pierde un poco de información, pero es una buena forma de tratar esta variable como continua.

**EDAD5:** muestra la edad del encuestado en rangos de 5 años, siendo el número que aparece el primero del rango. Por ejemplo, para el rango de 40 a 44 años, aparece el número 40.

**SEXO1:** Hace referencia al sexo del encuestado y su codificación es la siguiente:  
1=Hombre; 0=Mujer

**ECIV1:** Muestra el estado civil del encuestado y se divide entre solteros/as y casados/as de esta forma: 1=No Casado/a; 0=casado/a

**NAC1:** Nos indica la nacionalidad del encuestado con la siguiente codificación:  
1=tiene nacionalidad española (se incluye la doble nacionalidad); 0= no tiene nacionalidad española.

**NFORMA:** Indica la formación académica del encuestado y ha sido codificada de este modo:

0=sin estudios (analfabetos y primaria no acabada)

1=educación primaria completada

2=educación secundaria primera etapa

3=educación secundaria segunda etapa

4=estudios superiores<sup>1</sup>

## 2.2 Realización del análisis descriptivo y resultados

Para comenzar nuestro estudio lo primero que haremos será un análisis descriptivo de nuestra muestra, con ello pretendemos observar que porcentaje de desempleo hay en la muestra dependiendo de cada variable, como por ejemplo ver la diferencia de desempleo entre hombres y mujeres o entre cada nivel de estudios. También comprobaremos variables para ver como afectan juntas, por ejemplo, el desempleo entre hombres casados y solteros y mujeres casadas y solteras. Esto lo haremos con la aplicación *Statgraphics*.

---

<sup>1</sup> Nota: esta codificación es usada para el análisis descriptivo, a la hora de realizar el análisis logit se codificará esta variable de otra forma, añadiendo variables ficticias.

Antes de realizar el análisis de cada variable, veremos cual es la tasa de paro global de la muestra, para poder realizar comparaciones en torno a ella.

En el análisis realizado hemos obtenido una tasa de desempleo general del 20,96%.

### 2.2.1 Análisis de la variable “edad”

La primera variable que vamos a analizar es la edad, y lo haremos con la variable que ofrece los datos por grupos de 5 años. Se muestra a continuación un gráfico donde se ve claramente la tendencia del desempleo según la edad de los individuos.

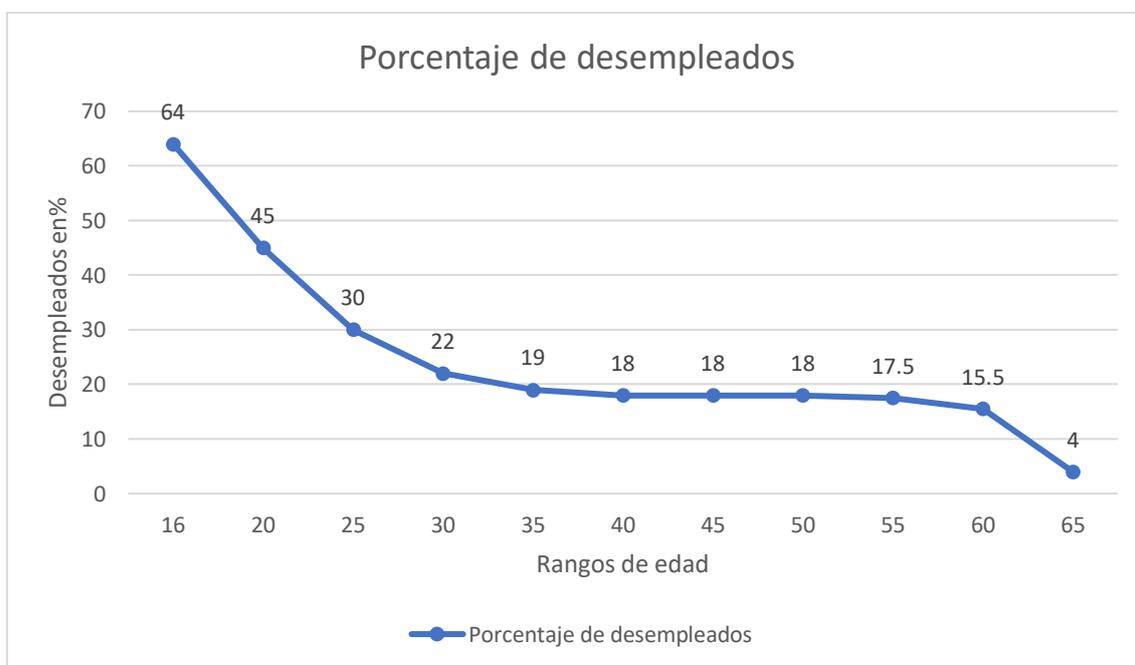


Gráfico 2. 1 Porcentaje de desempleados según la edad

En el gráfico podemos apreciar claramente como el porcentaje de desempleados muestra una tendencia descendente, la cual se estabiliza a partir de los 30 años.

Las conclusiones que sacamos de este análisis son dos: la primera, que el porcentaje de ocupados se muestra estable cuando el individuo se encuentra en un rango de mediana edad, entre los 30 y 60 años. Además de ver que conforme aumenta la edad las posibilidades de desempleo se reducen.

La segunda conclusión es que los jóvenes tienen muchas dificultades para encontrar empleo. Solo el 36% de los jóvenes de entre 16 y 20 años que quieren trabajar lo consiguen, porcentaje que va incrementándose según avanza la edad del encuestado, pero que no se sitúa en cifras más o menos normales hasta los 25 años, donde el porcentaje de ocupados alcanza el 70%, recordemos que la tasa general de paro es de casi un 21%, y en los individuos de 25 años aún es bastante superior.

### 2.2.2 Análisis de la variable “estado civil”

Analizaremos ahora la relación entre el estar ocupado o no con el estado civil del individuo, para comprobar si hay alguna diferencia significativa entre las personas solteras y casadas.

	Casado/a		No Casado/a	
	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta	Frecuencia relativa
0=Ocupado	34858	0,8418	23502	0,7246
1=Desempleado	6553	0,1582	8925	0,2752

*Tabla 2. 1 análisis descriptivo de la variable "estado civil"*

La diferencia en este caso es bastante significativa, siendo el porcentaje de desempleo casi 12 puntos mayor en el caso de las personas solteras. Como curiosidad, podemos observar, que aunque en la muestra haya casi nueve mil personas casadas más que solteras, el número total de desempleados solteros es mayor que el de casados.

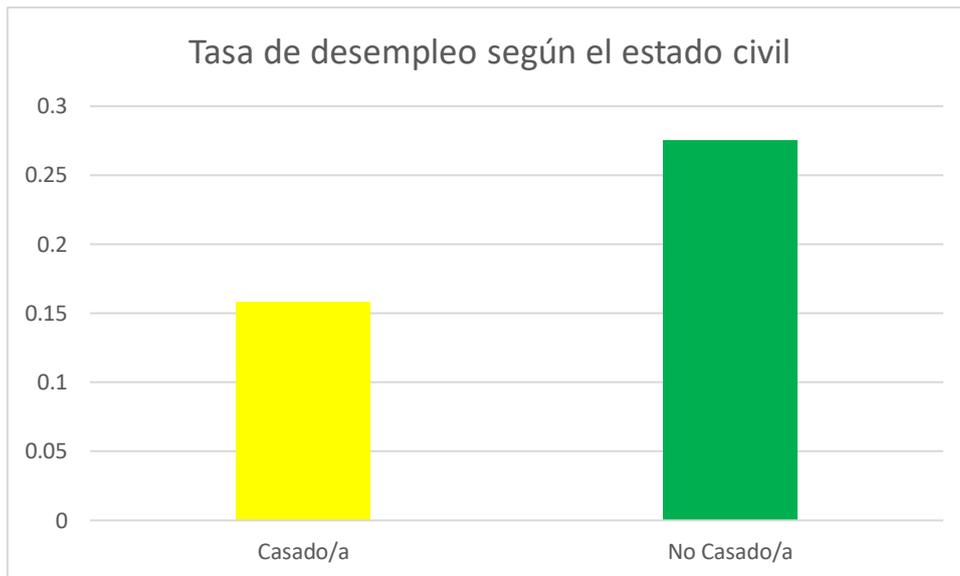


Gráfico 2. 2 Tasa de desempleo según el estado civil

### 2.2.3 Análisis de la variable “sexo”

A continuación, analizaremos la variable sexo y comprobaremos si hay diferencias significativas entre los hombres y las mujeres.

	Hombres		Mujeres	
	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta	Frecuencia relativa
0=Ocupado	31274	0,8061	27086	0,7730
1=Desempleado	7525	0,1939	7953	0,2270

Tabla 2. 2 Análisis descriptivo de la variable "sexo"

La diferencia en el desempleo entre hombres y mujeres es de unos 3 puntos porcentuales en favor de los hombres, variación suficiente como para poder afirmar que es una diferencia significativa y que el sexo influye en la facilidad para encontrar un empleo.

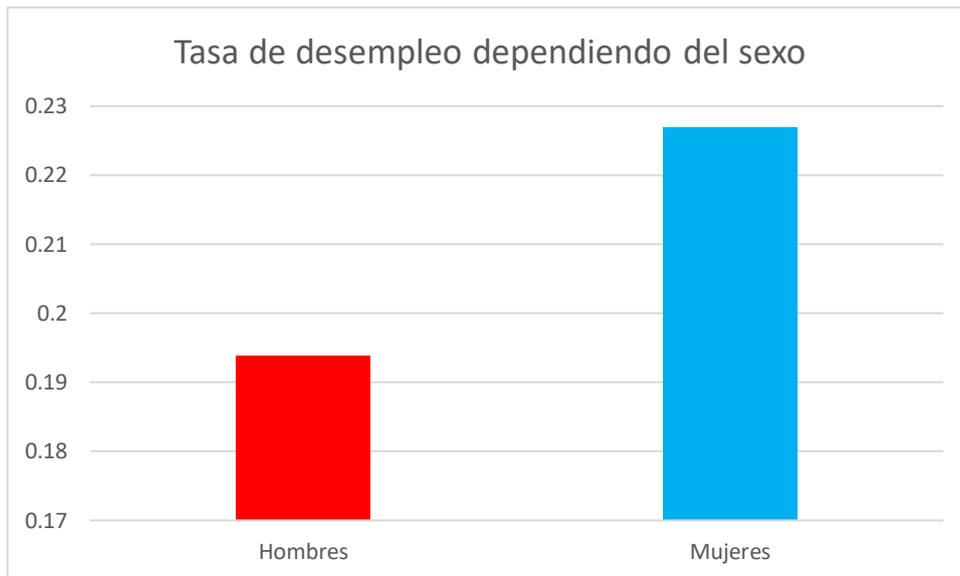


Gráfico 2. 3 Tasa de desempleo según el sexo

En este gráfico se puede apreciar esa diferencia entre ambos sexos. Ahora compararemos ambos sexos y su estado civil, en busca de posibles diferencias, ya que, a priori las mujeres casadas deberían tener una mayor tasa de desempleo que los hombres y que las mujeres no casadas, mientras que las mujeres no casadas deberían tener una tasa similar a los hombres. Veamos si es cierto.

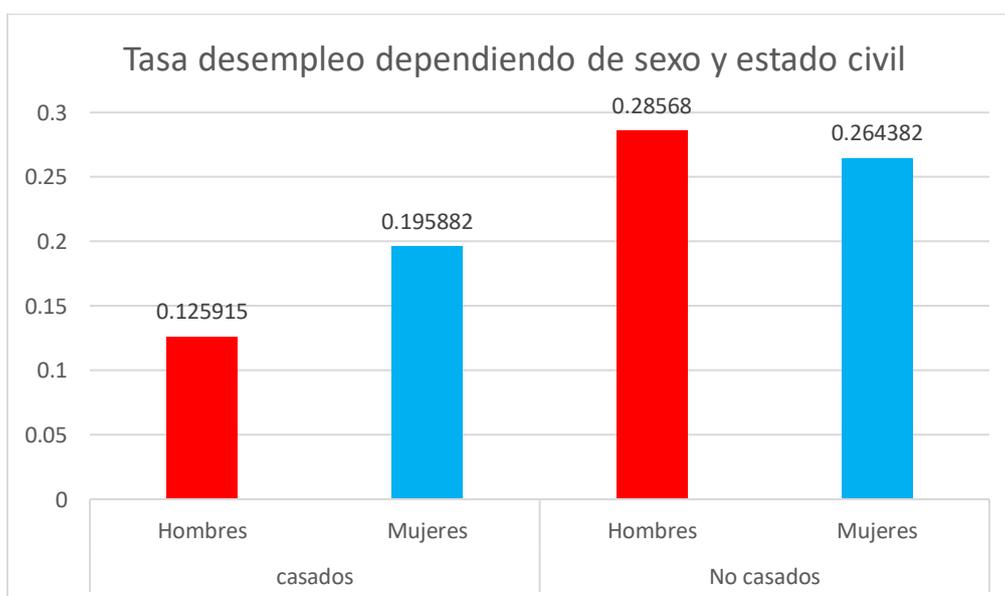


Gráfico 2. 4 Tasa de desempleo según el sexo y el estado civil

Vemos como en el caso de las personas casadas el desempleo en mujeres es muy superior, como ya se esperaba, mientras que entre las personas no casadas, no solo se reduce esa diferencia, sino que hay más tasa de desempleo entre los hombres.

#### 2.2.4 Análisis de la variable “nacionalidad”

Procedemos a realizar el análisis sobre la variable nacionalidad, para ver si el tener la nacionalidad española radica en un mayor porcentaje de ocupados.

	Posee nacionalidad española		Extranjeros/as	
	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta	Frecuencia relativa
0=Ocupado	54862	0,7976	3498	0,6924
1=Desempleado	13924	0,2024	1554	0,3076

*Tabla 2. 3 Análisis descriptivo de la variable "nacionalidad"*

Como vemos el número de observaciones de las personas extranjeras es mucho menor. A la hora de ver el número de ocupados nos encontramos con prácticamente un 70% de ocupados sin nacionalidad española, y, aunque el porcentaje de ocupados que sí cuentan con la nacionalidad española sea un 10% superior, el resultado de los extranjeros ha resultado gratamente sorprendente. Por lo cual, podemos concluir que la nacionalidad sí que influye en la facilidad para encontrar empleo, pero menos de lo esperado.

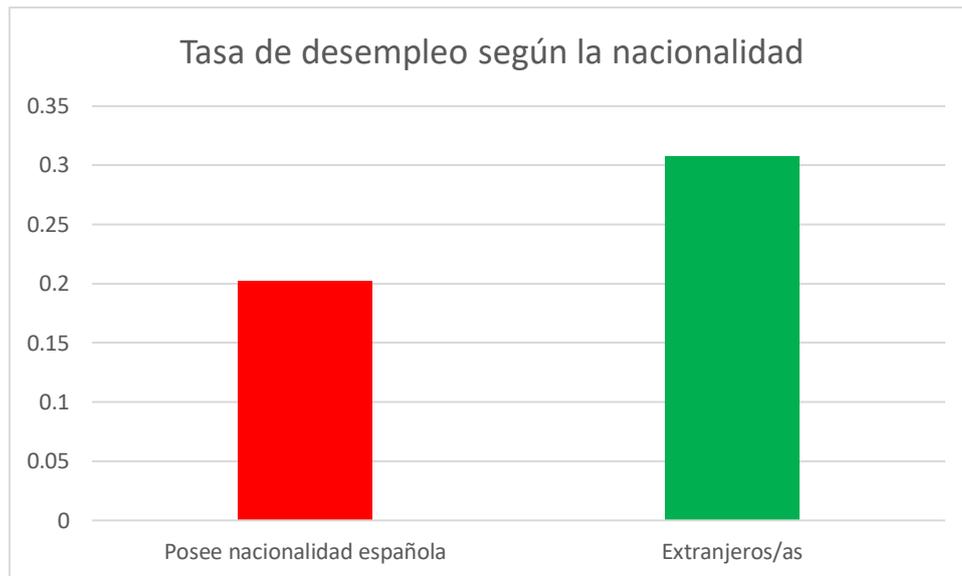


Gráfico 2. 5 Tasa de desempleo según la nacionalidad

### 2.2.5 Análisis de la variable “formación”

Por último, analizaremos la variable formación, y veremos si el hecho de tener unos estudios mayores ayuda a la hora de encontrar un empleo. Veremos primero a través de una tabla los porcentajes de ocupados y desempleados de las distintas categorías.

	sin estudios	primaria	secundaria1	secundaria2	superiores
ocupado	0,5916	0,654	0,7216	0,7968	0,8752
desempleado	0,4084	0,346	0,2784	0,2032	0,1248

Tabla 2. 4 Análisis descriptivo de la variable "formación"

En esta tabla lo primero que podemos ver es cómo según aumenta el nivel de estudios la diferencia entre ocupados y desempleados se hace cada vez más grande, partiendo de un 40% de desempleo en las personas sin estudios y llegando a un 12,5% de desempleo en las personas con estudios superiores.

A continuación mostramos un gráfico donde se puede apreciar mejor esa disminución del desempleo.

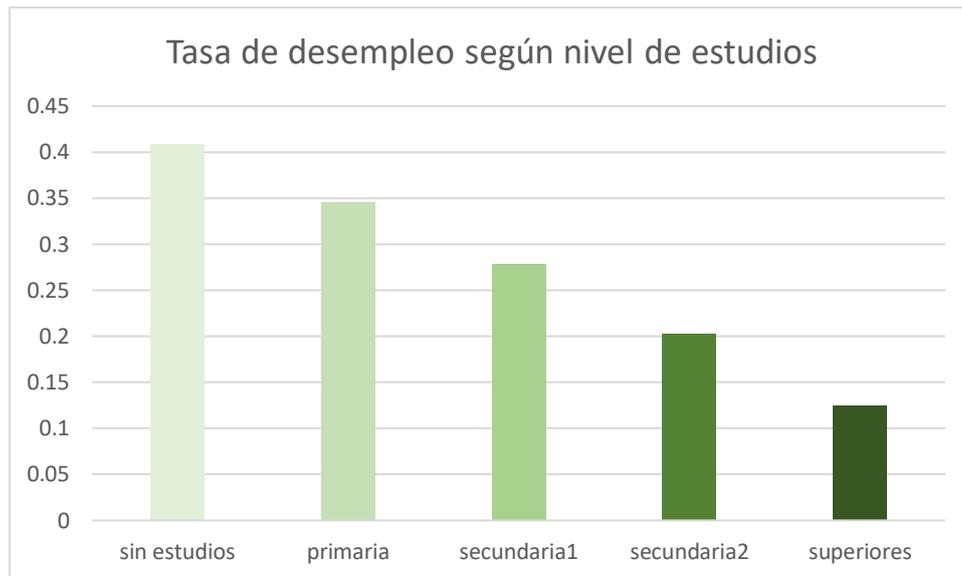


Gráfico 2. 6 Tasa de desempleo según el nivel de estudios

En las personas sin estudios y con educación primaria contamos con muy pocas observaciones, algo normal ya que en nuestro país es obligatorio completar hasta la primera parte de la educación secundaria, o hasta cumplir los 16 años. En estas dos categorías tenemos un desempleo de un 40% y un 34% respectivamente. Tasas de desempleo muy elevadas y alejadas de la media general, algo habitual ya que son personas muy poco cualificadas y solo pueden acceder a un número limitado de empleos.

Entre las personas con educación secundaria de primera etapa concluida el nivel de desempleo sigue su tendencia y se reduce hasta un 27%, cifra que se va aproximando a la tasa generales. Este nivel de estudios es el actualmente obligatorio en el país, por lo que tampoco son empleados altamente cualificados.

Entre las personas que han concluido la segunda etapa de educación secundaria el desempleo cae hasta un 20%, coincidiendo con la media general del desempleo. Aquí ya tenemos trabajadores con un nivel de cualificación más importante, incluyendo estudios de formación profesional o grado medio, los cuales ya otorgan unos conocimientos suficientes para poder optar a puestos de trabajo superiores que las categorías anteriormente mencionadas.

Por último, entre las personas con estudios superiores el desempleo se sitúa en su nivel más bajo, con un 12,5%. Una cifra lógica ya que son los trabajadores

más cualificados de la encuesta y hay muchos puestos de trabajo a los que sólo este grupo de encuestados puede optar.

Por lo tanto, como era de esperar, podemos concluir que un mayor nivel de estudios aumenta las posibilidades de conseguir trabajo. Además vemos una importante diferencia entre estudios superiores y la segunda etapa de educación secundaria, un tema muy recurrente últimamente debido a la calidad y cantidad de estudios no universitarios que existen. Por lo que mucha gente considera mejor opción que ir a la universidad el completar estudios de formación profesional y poder incorporarse antes al mercado laboral y con una buena cualificación, aunque en los datos podemos ver como las personas con estudios universitarios tienen más fácil la labor de encontrar empleo.

### 3. MODELO LOGIT

Ahora que ya tenemos la muestra reducida y hemos realizado un análisis descriptivo de la misma es momento de realizar el análisis logit, con la intención de conseguir una ecuación que la probabilidad de estar desempleado en nuestro país en función de unas determinadas variables, pero antes explicaremos unos fundamentos teóricos básicos sobre este modelo logit.

#### 3.1 Fundamentos teóricos

El modelo logit que empleamos en este trabajo es un modelo de elección discreta. Estos modelos se caracterizan por tratar de explicar una variable cualitativa, en nuestro caso el desempleo, en función de otras variables, ya sean cualitativas o cuantitativas. La variable dependiente del modelo sería:

$Y_i$	1	el individuo está desempleado
	0	el individuo está ocupado

Las variables independientes serán la edad, el estado civil, el sexo, el nivel de estudios y la nacionalidad.

La opción más sencilla dentro de los modelos de elección discreta sería el modelo lineal de probabilidad, pero en ese caso los valores estimados podrían ser superiores a 1, lo cual se soluciona usando una función de distribución que obligatoriamente toma sus valores en el intervalo [0,1] de esta manera:

$$Y_i = F(\beta_0 + \beta_1 \cdot X_{1i} + \beta_2 \cdot X_{2i} + \dots + \beta_k \cdot X_{ki}) + \varepsilon_i \quad i = 1, 2, \dots, N$$

Donde  $0 \leq F(.) \leq 1$

Por lo tanto

$$E(Y_i) = p(Y_i = 1) = p_i = F(X_i' \beta)$$

$$p(Y_i = 0) = (1 - p_i) = 1 - F(X_i' \beta)$$

El modelo logit asume que  $F$  es una función de distribución logística, donde la esperanza es cero y la desviación típica es  $\pi / \sqrt{3}$

Y la función quedaría así:

$$Y_i = \frac{e^{\beta_0 + \beta_1 \cdot X_{1i} + \beta_2 \cdot X_{2i} + \dots + \beta_k \cdot X_{ki}}}{1 + e^{\beta_0 + \beta_1 \cdot X_{1i} + \beta_2 \cdot X_{2i} + \dots + \beta_k \cdot X_{ki}}} + \varepsilon_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot X_{1i} + \beta_2 \cdot X_{2i} + \dots + \beta_k \cdot X_{ki})}} + \varepsilon_i \quad i = 1 \dots n$$

Un elemento importante en el modelo es el *Odd*, que es el cociente entre la probabilidad de que  $Y_i=1$  y la probabilidad de que  $Y_i=0$ . Si para un individuo este *Odd* es mayor que la unidad, significará que es más probable que  $Y_i=1$  que  $Y_i=0$ .

Otro concepto similar pero que no debemos confundir es el *Odd Ratio*. Es un cociente entre *Odds*. Si calculamos el *Odd Ratio* entre el *Odd* actual y el *Odd* habiendo aumentado una unidad de la variable manteniendo todos los valores constantes es  $e^{\beta_j}$  que tenemos en la tabla de "Variables de la ecuación" en la columna llamada "Exp(B)".

### 3.2. Análisis logit.

Para realizar el análisis logit hemos cambiado ligeramente las variables, para que el programa pueda arrojarnos un resultado sin errores. Lo primero que hemos hecho ha sido redefinir la variable del nivel de estudios. En el análisis descriptivo inicial teníamos una variable con valores de 0 a 4. Ahora lo que hemos hecho ha sido introducir n-1 variables ficticias. “PRIM”, “SEC1”, “SEC2” y “SUP” que entregan valores de 1 si el individuo tiene ese nivel de estudios y 0 si no ha alcanzado ese nivel o si tiene uno superior. Estas variables corresponden a educación primaria, educación secundaria primera etapa, educación secundaria segunda etapa y estudios superiores respectivamente.

El otro cambio realizado ha sido en la variable edad. En el análisis descriptivo se utilizó la variable “EDAD5” que separaba a los individuos en grupos de edad quinquenales, ahora usaremos “EDADcont” para tratar a la edad como una variable continua.

Una vez hechos estos cambios, mediante el programa “SPSS” se ejecutó un análisis logit que arrojó los siguientes resultados.

		Variables en la ecuación					
		B	Error estándar	Wald	gl	Sig.	Exp(B)
Paso 1 <sup>a</sup>	EDADcont	-,028	,001	939,000	1	<,001	,972
	SEX01	-,333	,019	303,774	1	<,001	,717
	ECIV1	,488	,021	536,566	1	<,001	1,629
	NAC1	-,249	,034	52,325	1	<,001	,780
	PRIM	-,332	,069	23,375	1	<,001	,718
	SEC1	-,770	,063	147,714	1	<,001	,463
	SEC2	-1,293	,065	400,553	1	<,001	,275
	SUP	-,471	,016	848,985	1	<,001	,625
	Constante	1,195	,080	221,083	1	<,001	3,305

a. Variables especificadas en el paso 1: EDADcont, SEX01, ECIV1, NAC1, PRIM, SEC1, SEC2, SUP.

Tabla 3. 1 Variables en la ecuación modelo logit

En esta tabla tenemos los datos para sacar la ecuación del logit. En la columna “B” vemos unos datos de los que solo interpretaremos el signo. Por ejemplo, en la variable “EDADcont” tenemos signo negativo. Esto significa que conforme aumenta esa variable las probabilidades de estar desempleado disminuyen. Pero no disminuyen en ese valor, solo podemos afirmar que disminuyen, para ver cuánto disminuyen haremos otra prueba después para ver el efecto marginal de cada variable.

Además de esos datos, podemos ver en la columna “Sig.” Que todos los valores son menores a 0,05, por lo que todas las variables son significativas.

Y por último tenemos la columna “Exp(B)” que representa el *Odd Ratio*. El *Odd Ratio* representa la variación de *Odd* cuando la variable aumenta una unidad y el resto permanece constante.

*Odd* es el cociente entre la probabilidad de que  $Y=1$  y  $Y=0$ .

Volviendo al ejemplo de la edad. Cuando la edad aumenta en una unidad, el *Odd* se multiplicará por 0,972. Es decir, el cociente entre que  $Y=1$  y  $Y=0$  disminuirá en 2,8%

Después de este análisis esta es la ecuación que hemos obtenido para el modelo logit:

$$\frac{1}{1 + e^{-(1,195 - 0,028 \cdot X_{1i} - 0,333 \cdot X_{2i} + 0,488 \cdot X_{3i} - 0,249 \cdot X_{4i} - 0,332 \cdot X_{5i} - 0,770 \cdot X_{6i} - 1,293 \cdot X_{7i} - 0,471 \cdot X_{8i})}}$$

A la hora de analizar los aciertos en el pronóstico de nuestro modelo nos encontramos con el siguiente problema

**Tabla de clasificación<sup>a</sup>**

Observado		Pronosticado		Porcentaje correcto
		AOIB 0	AOIB 1	
Paso 1	AOIB 0	57488	872	98,5
	AOIB 1	14460	1018	6,6
Porcentaje global				79,2

a. El valor de corte es ,500

Tabla 3. 2 Porcentaje de aciertos inicial modelo logit

Como podemos ver, el porcentaje de aciertos es muy alto entre las personas con un 0 en la variable AOIB, es decir, las personas ocupadas. Pero entre las personas desempleadas el porcentaje de aciertos es de solo un 6,6%. Este modelo tiende a clasificar las personas como ocupadas.

El motivo de que esto ocurra es que el análisis logit otorga una probabilidad de estar desempleado de entre 0 y 1, asignando al individuo al grupo de ocupados o desempleados según esta probabilidad, de modo que si es mayor de 0,5 asignará al individuo al grupo de desempleados y si es menor al de ocupados, de modo que casi todos tenían una probabilidad menor de 0,5.

Esto no es muy realista, puesto que la tasa de paro general de la muestra es de aproximadamente un 0,21 y no 0,5, de modo que cambiando este parámetro del análisis el programa pronosticará como desempleados a los individuos con una probabilidad de desempleo superior a 0,21, lo cual es lo correcto. Vemos que tras este cambio el porcentaje de aciertos varía bastante.

**Tabla de clasificación<sup>a</sup>**

Observado		Pronosticado		Porcentaje correcto
		AOIB 0	AOIB 1	
Paso 1	AOIB 0	38195	20165	65,4
	AOIB 1	5785	9693	62,6
Porcentaje global				64,9

a. El valor de corte es ,210

*Tabla 3. 3 Porcentaje de aciertos final modelo logit*

Como vemos ha mejorado bastante y ya no clasifica a casi todos como ocupados, además el porcentaje de aciertos entre ocupados y desempleados es bastante similar, lo cual difiere mucho del modelo inicial y nos indica que el cambio que hemos realizado ha sido efectivo.

### 3.2.1 Efecto de cada variable

Ahora veremos el efecto que cada variable tiene de forma individual sobre el modelo, es decir, manteniendo el resto de variables constantes. En este modelo tenemos dos tipos de variables: continuas, que en nuestro caso es solo la variable “edad” y dicotómicas, como son el resto de variables. Para ver el efecto de la variable “edad” calcularemos su efecto marginal mediante el programa SPSS usando la siguiente fórmula:  $P_i*(1-P_i)*\beta$ , donde “ $P_i$ ” es la probabilidad que nos pronosticaba el programa de estar desempleado para cada individuo y  $\beta$  es el valor de la variable edad en el modelo, en nuestro caso -0,028. Obtendremos así un resultado para cada individuo, y haciendo la media vemos el efecto marginal de esta variable, que en nuestro caso es de 0,0043. Es decir, cuando la edad aumenta en una pequeña porción, las probabilidades de estar desempleado se reducen en 0,0043 puntos porcentuales siempre que el resto de variables se mantenga constante. En cambio, para analizar el efecto de las variables dicotómicas lo haremos de otra forma, ya que, el estado civil, por ejemplo, no puede aumentar en una unidad, el individuo solo puede estar casado o no estarlo. Lo que haremos para estas variables será calcular la probabilidad de estar desempleado siendo la variable 1 y después calcularlo siendo 0, manteniendo el resto de variables constantes en su valor medio. En el caso de los estudios, que tenemos 4 variables ficticias, calcularemos la probabilidad de estar desempleado teniendo un 1 en cada una mientras hay un 0 en las demás y las otras variables referentes al sexo, estado civil, nacionalidad y edad se mantienen en su valor medio. Se muestra a continuación una tabla con el efecto de cada variable:

	Respecto al nivel anterior	Respecto a sin estudios
Efecto de la variable educación primaria	8%	
Efecto de la variable educación secundaria primera etapa	9,5%	17,5%
Efecto de la variable educación secundaria segunda etapa	9%	26,5%
Efecto de la variable estudios superiores	7%	33,5%

Tabla 3. 4 Efecto de las variables referentes al nivel de estudios

Esto quiere decir que el tener la educación primaria reduce en un 8% las probabilidades de estar desempleado respecto a los individuos sin estudios. El tener completada la primera etapa de educación secundaria reduce la probabilidad de estar desempleado un 9,5% respecto a los individuos con educación primaria y un 17,5% respecto a los individuos sin estudios, y así sucesivamente. Podemos apreciar una gran diferencia conforme el nivel de estudios aumenta, llegando a ser un 33,5% más probable encontrar empleo para una persona con estudios superiores que para otra con sus mismas características, pero sin estudios.

Variable	Efecto de la variable (puntos porcentuales)	Probabilidad de desempleo con un 1 en la variable	Probabilidad de desempleo con un 0 en la variable
Sexo	5	16,1%	21,1%
Estado civil	7,5	22,8%	15,3%
Nacionalidad	3,98	18,07%	22,05%

*Tabla 3. 5 Efecto de las variables "sexo" "estado civil" y "nacionalidad"*

En el caso de estas variables, es un 5% más probable encontrar trabajo siendo hombre, un 7,5% más probable encontrar trabajo estando casado y un 4% aproximadamente más probable encontrar trabajo si tienes la nacionalidad española, siempre que el resto de variables se mantengan constantes. Son resultados un tanto sorprendentes, ya que antes de empezar a realizar este trabajo esperaba ver un gran efecto de la variable nacionalidad y ha resultado ser la variable menos determinante, mientras que una variable como el estado civil tiene una relevancia que para nada esperaba.

#### **4. Escenarios**

En este apartado vamos a elegir unos escenarios para ver las diferencias en las probabilidades de encontrar trabajo. Elegiremos distintos individuos variando

algunas de sus características para poderlos comparar por parejas o pequeños grupos y ver como afectan ciertas características a individuos concretos.

Edad	Sexo	Estado Civil	Nacionalidad	Formación	Probabilidad pronosticada
20	Hombre	No casado/a	Española	secundaria 1	0,4435
20	Mujer	No casado/a	Española	secundaria 1	0,52641
20	Hombre	No casado/a	Extranjera	secundaria 1	0,50545
20	Mujer	No casado/a	Extranjera	secundaria 1	0,58772
25	Hombre	No casado/a	Española	superiores	0,18544
25	Mujer	No casado/a	Española	superiores	0,241
25	Hombre	No casado/a	Extranjera	superiores	0,22598
30	Hombre	Casado/a	Española	secundaria 2	0,17991
30	Mujer	Casado/a	Española	secundaria 2	0,23429
30	Hombre	Casado/a	Extranjera	secundaria 2	0,21957
30	Mujer	Casado/a	Extranjera	secundaria 2	0,28182
30	Hombre	No casado/a	Española	secundaria 2	0,26326
30	Mujer	No casado/a	Española	secundaria 2	0,33262
35	Hombre	Casado/a	Española	superiores	0,09559
35	Mujer	Casado/a	Española	superiores	0,12848
35	Mujer	No casado/a	Española	superiores	0,19363
45	Hombre	Casado/a	Española	secundaria 2	0,12608
45	Mujer	Casado/a	Española	secundaria 2	0,16752
45	Hombre	Casado/a	Española	superiores	0,07402
45	Mujer	Casado/a	Española	superiores	0,10031

48	Hombre	Casado/a	Española	superiores	0,18285
48	Hombre	No casado/a	Española	secundaria 1	0,26712
48	Hombre	Casado/a	Extranjera	secundaria 1	0,22299
50	Hombre	Casado/a	Española	sin estudios	0,31367
50	Mujer	Casado/a	Española	sin estudios	0,38929
60	Hombre	No casado/a	Española	primaria	0,28778
60	Mujer	No casado/a	Española	primaria	0,36045
60	Hombre	Casado/a	Española	primaria	0,19876
60	Mujer	Casado/a	Española	primaria	0,25706

Tabla 4. 1 Distintos escenarios

Los primeros cuatro individuos son jóvenes de 20 años, con pequeñas diferencias entre sí. Hay dos hombres y dos mujeres, siendo españoles un hombre y una mujer y extranjeros los otros dos restantes. Todos tienen el mismo nivel de estudios, que son los estudios secundarios de primera etapa, y ninguno está casado. Podemos ver que las probabilidades de estar desempleados son muy altas para todos debido a su corta edad. El hombre español tiene la menor tasa de desempleo, mientras que la mujer extranjera es el individuo con más probabilidades de estar en paro. Entre medias el hombre extranjero tiene dos puntos porcentuales menos en su probabilidad de estar desempleado que la mujer española.

En el siguiente grupo de edad de 25 años tenemos tres individuos, una mujer española y dos hombres, uno español y el otro extranjero, los tres con estudios superiores y sin estar casados. En este caso se repite el orden del grupo anterior, siendo el hombre español el que menos posibilidades de desempleo tiene y la mujer española la que más, aunque el aumento de la edad y el nivel de estudios hacen que las tres probabilidades se encuentren cercanas a la tasa general de desempleo, el hombre español trabajaría, con una probabilidad de un 10,5%, mientras que los otros dos individuos están en un 22% y un 24% aproximadamente.

Posteriormente tenemos seis individuos de 30 años y estudios secundarios de segunda etapa. Dentro de esos seis individuos tenemos tres hombres y tres mujeres. Una pareja de casados españoles, otra de casados extranjeros y una última de solteros con nacionalidad española. Los hombres casados son los que menor probabilidad de desempleo tienen, siendo la del español bastante menor, con un 18% y un 22% aproximadamente. A continuación la mujer española casada, con un pronóstico de un 23% de probabilidad de desempleo, seguida del hombre soltero, con un 26%. La mujer extranjera tiene un 28% de probabilidades de estar desempleada, y por último la mujer soltera, cuyas probabilidades de estar desempleada son de más de un 33%.

Posteriormente tenemos tres individuos de 35 años con estudios superiores y nacionalidad española. Un hombre casado y dos mujeres, una casada y la otra no. El hombre vuelve a tener el mejor pronóstico, con solo un 9% de probabilidades de desempleo, mientras que entre las mujeres la diferencia en el estado civil se ve reflejada en su pronóstico laboral, siendo mucho más difícil encontrar trabajo para la mujer soltera.

A continuación tenemos cuatro individuos de 45 años, dos hombres y dos mujeres, todos casados y españoles. Entre ellos un hombre y una mujer tienen estudios superiores y los otros dos estudios secundarios de segunda etapa. En este caso los dos individuos con estudios superiores tienen las menores posibilidades de estar desempleados, siendo mejor el pronóstico del hombre. Los otros dos individuos con un nivel menos de estudios no se alejan mucho, también tienen unas probabilidades por debajo de la tasa general de desempleo.

A continuación tenemos tres individuos de 48 años, todos hombres pero con distintas características. Uno está casado, es español y tiene estudios superiores. Los otros dos solo tienen completada la primera etapa de educación secundaria, siendo uno de ellos español y soltero, mientras que el otro es extranjero y casado. El que mejor pronóstico tiene es el primero de ellos, algo que era de suponer, pero entre los otros dos tiene un pronóstico más favorable el extranjero, gracias a su estado civil.

Comentamos ahora dos personas de 50 años, hombre y mujer, ambos españoles, casados y sin estudios. El pronóstico no es muy bueno debido al

pobre nivel de estudios que tienen y de nuevo el hombre sale mejor parado en esta comparativa.

Por último tenemos un grupo de cuatro personas de 60 años. Dos hombres y dos mujeres, todos españoles y con estudios primarios. Un hombre y una mujer están casados mientras que los otros dos no. La falta de estudios se ve compensada con la edad y el hombre casado tiene un pronóstico favorable al empleo, con una probabilidad de estar desempleado de menos del 20%. Le sigue la mujer casada, pero a seis puntos de diferencia, mientras que el estado civil vuelve a jugar un papel fundamental en el pronóstico de los otros dos individuos, situándose muy lejos de la tasa general de desempleo, con unas probabilidades pronosticadas de casi un 29% para el y de un 36% para ella.

Como conclusión obtenemos que las mejores características para un individuo a la hora de buscar empleo serían las siguientes: un hombre, casado, con nacionalidad española, con estudios superiores y de avanzada edad, como hemos visto en anteriores apartados, la edad es inversamente proporcional a las posibilidades de desempleo. Por el contrario, el individuo más propenso a estar en desempleo sería una mujer, soltera, sin nacionalidad española ni estudios y con una edad joven.

Como hemos visto tanto en este apartado de escenarios como en algunos anteriores, el estado civil es la variable que más afecta al desempleo entre las tres variables dicotómicas (sexo, estado civil y nacionalidad) y la nacionalidad la que menos.

#### 4.1 Curvas de respuesta

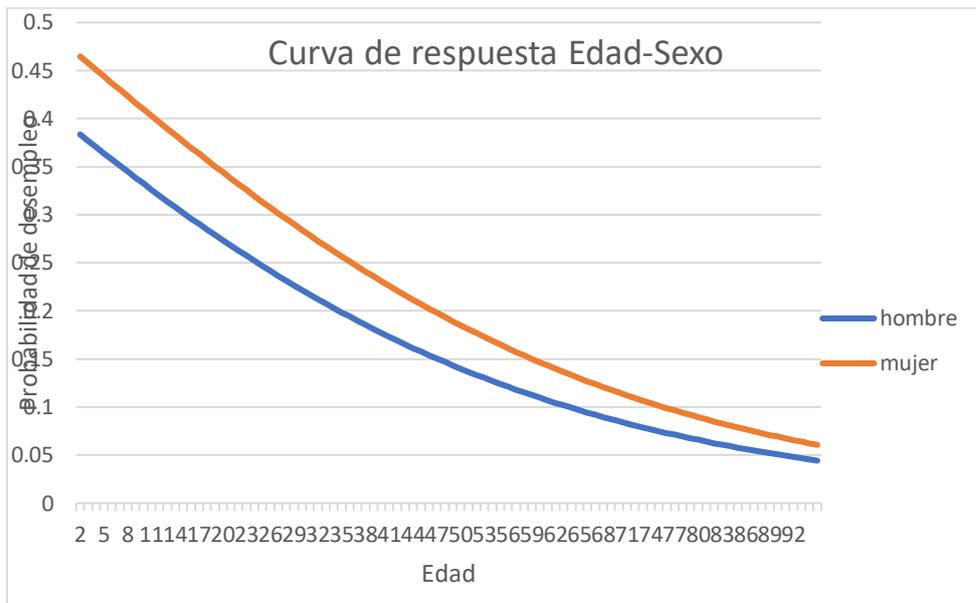


Gráfico 4. 1 Curva de respuesta edad-sexo

En este gráfico vemos las probabilidades de desempleo de ambos sexos a medida que avanza la edad. Siempre es menos probable el desempleo para los hombres, pero se aprecia claramente como la diferencia entre ambos sexos se reduce a medida que la edad avanza.

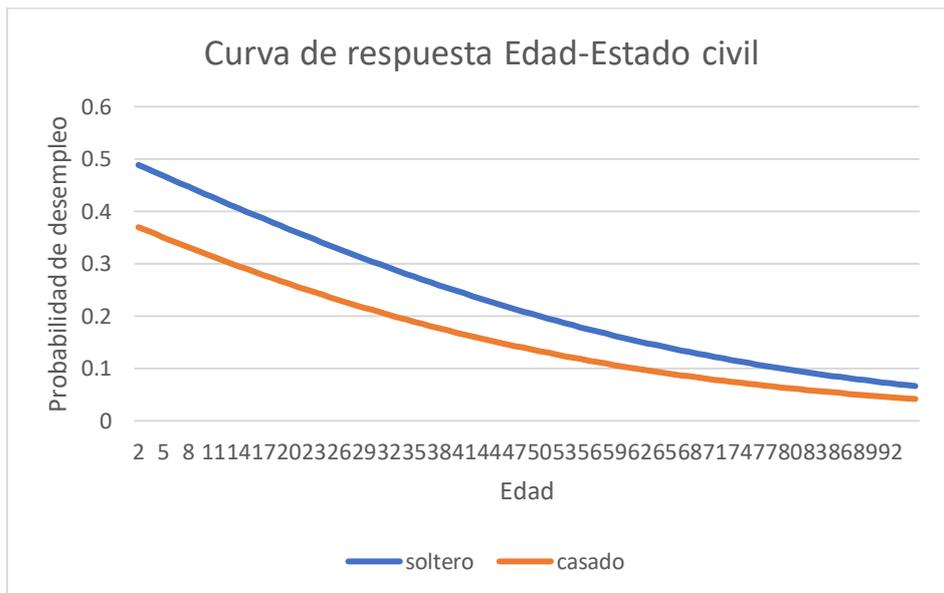


Gráfico 4. 2 Curva de respuesta edad-estado civil

Algo similar ocurre en el caso del estado civil. Siempre es más sencillo para los individuos casados encontrar trabajo, pero a medida que la edad aumenta la diferencia con las personas solteras se reduce.

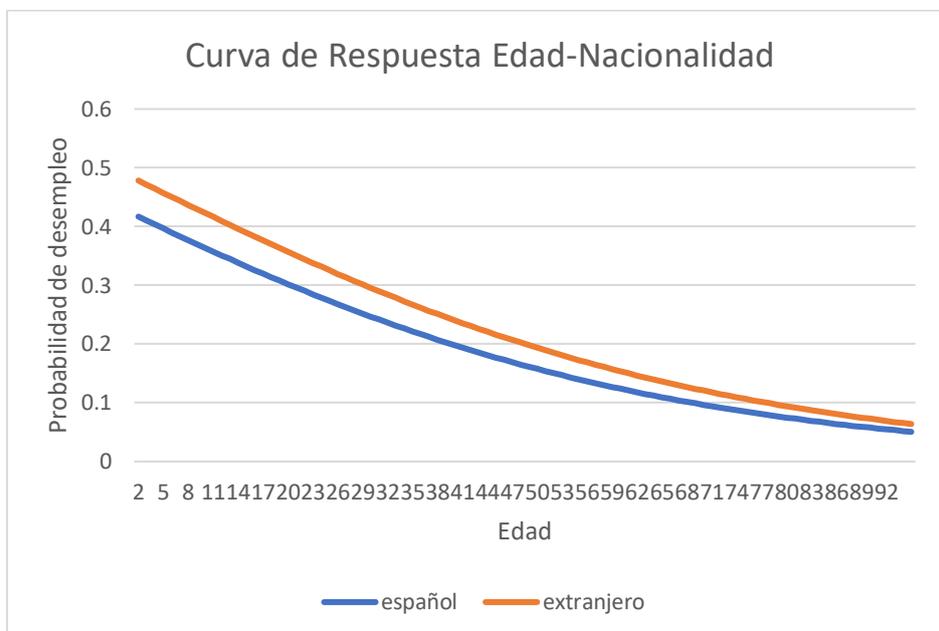


Gráfico 4. 3 Curva de respuesta edad-nacionalidad

Lo mismo ocurre en el caso de la nacionalidad. En este caso se puede apreciar como la diferencia es menor entre ambos grupos, ya que como hemos comentado anteriormente esta es la variable que menos influye en cuanto al empleo. En las edades más avanzadas se puede ver como prácticamente se tocan ambas líneas, teniendo tanto nacionales como extranjeros unas probabilidades de desempleo muy similares entre sí, pero siempre es más probable el desempleo para las personas que no poseen la nacionalidad española.

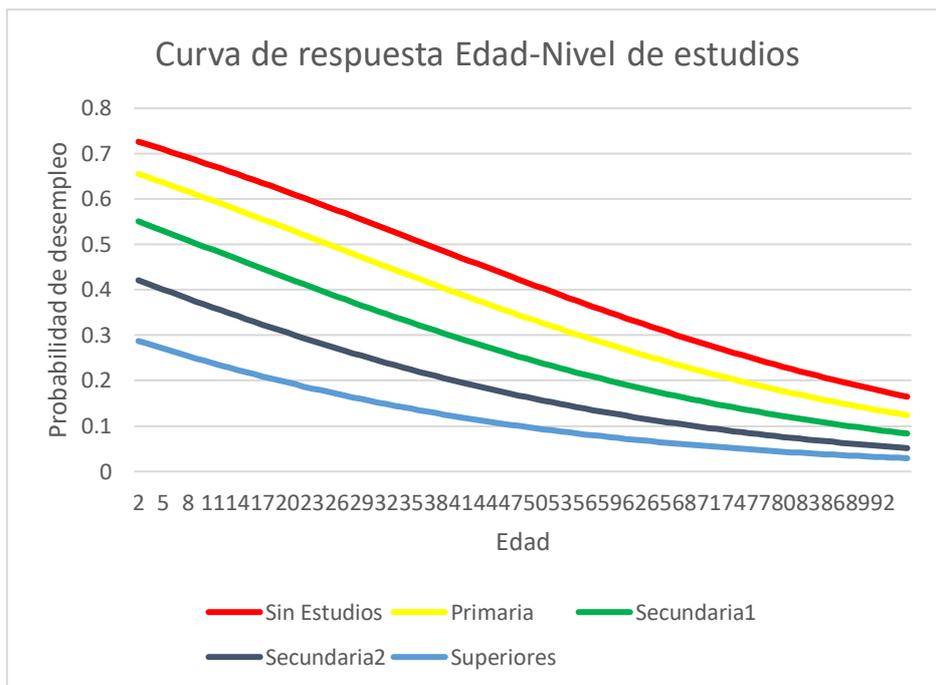


Gráfico 4. 4 Curva de respuesta edad-nivel de estudios

Por último tenemos la gráfica del nivel de estudios. Como es lógico a mayor nivel de estudios menores probabilidades de desempleo, y como ha ocurrido con las variables anteriores las diferencias se van reduciendo conforme aumenta la edad.

## 5. Conclusiones

Este trabajo ha consistido en la elaboración de un modelo logit para tratar de explicar los determinantes del desempleo en España. Primeramente se dio una explicación breve sobre el desempleo en España y por qué es algo tan importante, para continuar con un análisis descriptivo de la muestra seleccionada en la que se ha podido apreciar que grupos están más favorecidos y cuales menos a la hora de encontrar empleo. Por ejemplo, se puede ver cómo las mujeres tienen más dificultades que los hombres a la hora de encontrar empleo, teniendo 3 puntos porcentuales más de tasa de desempleo que los hombres. En cambio, si nos fijamos en el estado civil además de en el sexo, vemos como en el grupo de individuos no casados la tasa de paro femenina es de dos puntos menor que la masculina, mientras que entre los individuos casados la diferencia vista inicialmente entre sexos se incrementa, con una probabilidad de desempleo

mucho menor para los hombres, de aproximadamente 7 puntos. Si analizamos la variable estado civil únicamente observamos que es mucho más fácil encontrar empleo para las personas casadas, habiendo una diferencia de casi 8 puntos entre ambos grupos. Atendiendo a la nacionalidad del individuo vemos que el tener nacionalidad española hace más sencilla la tarea de encontrar trabajo que para los individuos no nacionales, con unas tasas de desempleo aproximadas de 0,2 y 0,3 respectivamente. Por último, en cuanto al nivel de estudios vemos como lógicamente a mayor nivel de estudios menor es la tasa de paro. Pasando de una tasa de paro superior al 40% para los individuos sin estudios a un escaso 12,5% de desempleo para los individuos con estudios superiores.

A continuación se presenta la parte del logit, que incluye unos pequeños conceptos teóricos sobre este modelo para seguir con el modelo en sí. El análisis se ha realizado mediante el programa *SPSS* y se ha conseguido un modelo logit mediante el cual pronosticar las probabilidades de desempleo de un individuo dependiendo de sus características. Gracias a este modelo hemos podido comprobar que la variable dicotómica que más afecta a la probabilidad de desempleo es el estado civil, ya que, manteniendo el resto de variables constantes, los individuos solteros tienen una probabilidad de desempleo de 7,5 puntos mayor que los casados. Por otra parte la variable que menos afecta al desempleo es la nacionalidad, con una diferencia inferior a los cuatro puntos entre ambos grupos, siendo la menor tasa de desempleo para los individuos con nacionalidad española.

A continuación se incluye un apartado de escenarios donde se han elegido diversos individuos y se han calculado sus probabilidades de desempleo mediante el modelo previamente mencionado. Aquí hemos podido ver la probabilidad de desempleo de ciertos individuos tipo. La menor probabilidad de desempleo corresponde a un individuo varón, casado, con nacionalidad española, estudios superiores y de avanzada edad, mientras que la mayor probabilidad de desempleo corresponde a una mujer joven sin nacionalidad española, soltera y sin estudios.

## 6. Bibliografía

*Comparar economía países: España vs Italia Tasa de desempleo (EPA) 2022.*

(s. f.). datosmacro.com.

<https://datosmacro.expansion.com/paises/comparar/espana/italia?sc=XE>

18

*Evolución de la tasa de paro en España desde 1975.* (s. f.). EpData.

<https://www.epdata.es/evolucion-tasa-paro-espana/2fc3fced-0264-4a55-bca8-d8c79509f4ae>

INE - Instituto Nacional de Estadística. (T12016). *INEbase / Mercado laboral*

*/Actividad, ocupación y paro /Encuesta de población activa /*

*Resultados [Conjunto de datos].*

[https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica\\_C&cid=1254736176918&menu=resultados&idp=1254735976595#!tabs-1254736030639](https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176918&menu=resultados&idp=1254735976595#!tabs-1254736030639)

Peterson, M. (2022, 1 marzo). *Por qué España tiene habitualmente una de las*

*tasas de desempleo más altas de Europa.* Foundation for Economic

Education. <https://fee.org.es/articulos/por-qu%C3%A9-espa%C3%B1a-tiene-habitualmente-una-de-las-tasas-de-desempleo-m%C3%A1s-altas-de-europa/>

*de-europa/*