



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 EL ANÁLISIS
DE LOS DETERMINANTES DEL
DESEMPLEO**

Presentado por:

Javier Colinas Mezquita

Valladolid, 10 de julio de 2023

RESUMEN

El desempleo es uno de los problemas estructurales más característicos de la economía española, contando nuestro país con una de las tasas de paro más altas de toda la Unión Europea.

El objetivo de este trabajo es la realización de un modelo Logit (apoyado en previos análisis descriptivos) para estudiar de manera cuantitativa el impacto de diferentes variables personales en la probabilidad de estar en situación de desempleo. Las variables utilizadas son la edad, el sexo, la nacionalidad, el estado civil y el nivel de estudios.

Los datos utilizados para el desarrollo de este estudio han sido recogidos del Instituto Nacional de Estadística (*INE*) y corresponden a la Encuesta de Población Activa (*EPA*) del primer trimestre de 2022.

Para los análisis estadísticos se ha utilizado el programa *SPSS* y *Excel* para el resumen de resultados y maquetación de tablas y el cálculo de la probabilidad estimada de diferentes escenarios con los resultados del modelo Logit.

Una vez realizado los análisis, podemos concluir en que es más probable encontrarse en situación de desempleo en mujeres, extranjeros, solteros y personas con bajo nivel de estudios, influyendo de manera muy positiva los estudios superiores a la hora de conseguir un empleo, con porcentajes significativamente más bajos en todos los tramos de edad, una variable que además afecta de manera positiva a la hora de conseguir trabajo, siendo más probable tener un empleo con una mayor edad.

Palabras clave: desempleo, Logit, SPSS

SUMMARY

Unemployment is one of the most characteristic structural problems of the Spanish economy, with our country having one of the highest unemployment rates in the entire European Union.

The objective of this study is to develop a Logit model (supported by previous descriptive analyses) to quantitatively examine the impact of different personal variables on the probability of being unemployed. The variables used are age, gender, nationality, marital status, and educational level.

The data used for this study were collected from the National Institute of Statistics (*INE*) and correspond to the Active Population Survey (*EPA*) for the first quarter of 2022.

The *SPSS* program was used for the statistical analysis, and *Excel* was used for summarizing the results, creating tables, and calculating the estimated probability of different scenarios using the Logit model results.

After conducting the analyses, we can conclude that it is more likely to be unemployed for women, foreigners, singles, and individuals with a lower educational level. Higher education has a very positive influence on obtaining employment, with significantly lower percentages across all age groups. Age is also a positive factor in obtaining employment, as it is more likely to have a job at an older age.

Keywords: unemployment, Logit, SPSS

ÍNDICE

1. INTRODUCCIÓN	1
2. ANÁLISIS DESCRIPTIVO DE LAS VARIABLES	5
2.1. Metodología	6
2.2. Variables	6
2.2.1. Situación laboral	6
2.2.2. Edad	8
2.2.3. Sexo.....	10
2.2.4. Nivel de estudios.....	12
2.2.5. Nacionalidad	14
2.2.6. Estado civil.....	16
3. MODELO LOGIT	17
3.1. Introducción teórica.....	17
3.2. Aplicación del modelo Logit.....	19
4. ESCENARIOS.....	24
5. CURVAS DE RESPUESTA.....	27
6. CONCLUSIONES	29
7. BIBLIOGRAFÍA	33
ANEXO.....	35

ÍNDICE DE GRÁFICOS

Gráfico 1 Evolución anual de la tasa de paro en España	1
Gráfico 2 Evolución de la tasa de desempleo en la UE y en la Eurozona	2
Gráfico 3 Porcentaje de desempleo por grupo de edad de la muestra	10
Gráfico 4 Porcentaje de desempleo por sexo de la muestra	11
Gráfico 5 Porcentaje de desempleo por nivel de estudios de la muestra	13
Gráfico 6 Porcentaje de desempleo por nacionalidad de la muestra	15
Gráfico 7 Porcentaje de desempleo por estado civil de la muestra	16
Gráfico 8 Curva de respuesta Edad - Nivel de estudios	27
Gráfico 9 Curva de respuesta Edad - Sexo	28
Gráfico 10 Curva de respuesta Edad - Nacionalidad	28
Gráfico 11 Curva de respuesta Edad - Estado civil	29

ÍNDICE DE TABLAS

Tabla 1 Frecuencias de la variable Situación laboral.	8
Tabla 2 Edad y Situación laboral.....	9
Tabla 3 Sexo y Situación laboral	11
Tabla 4 Nivel de estudios y Situación laboral.....	12
Tabla 5 Nivel de estudios, Sexo y Situación laboral.....	13
Tabla 6 Nacionalidad y Situación laboral	14
Tabla 7 Nacionalidad, Sexo y Situación laboral	15
Tabla 8 Estado civil y Situación laboral	16
Tabla 9 Estado civil, Sexo y Situación laboral	17
Tabla 10 Resultados del análisis Logit en SPSS.....	21
Tabla 11 Clasificación de los casos acorde al modelo Logit.	23
Tabla 12 Clasificación de los casos para $p(Y_i = 1) > 0,13$	23
Tabla 13 Escenarios.....	24

1. INTRODUCCIÓN

El desempleo es una de las principales cuestiones dentro de la economía española¹. Siendo este un problema estructural que se traduce en una alta tasa de paro natural y en consecuencia una tasa real muy por encima de la media de la Unión Europea y de la Eurozona².

La tasa de paro tras los resultados obtenidos en el último trimestre de 2022 según el Instituto Nacional de Estadística (INE en adelante) se sitúa en un 12,87%³, un dato muy por encima del respectivo a la Unión Europea (6,1%) y a la eurozona (6,7%).

Sin embargo, a pesar de duplicar la tasa de paro de la media europea, este no es del todo un dato desesperanzador, pues es la tasa de desempleo más baja desde el año 2007.

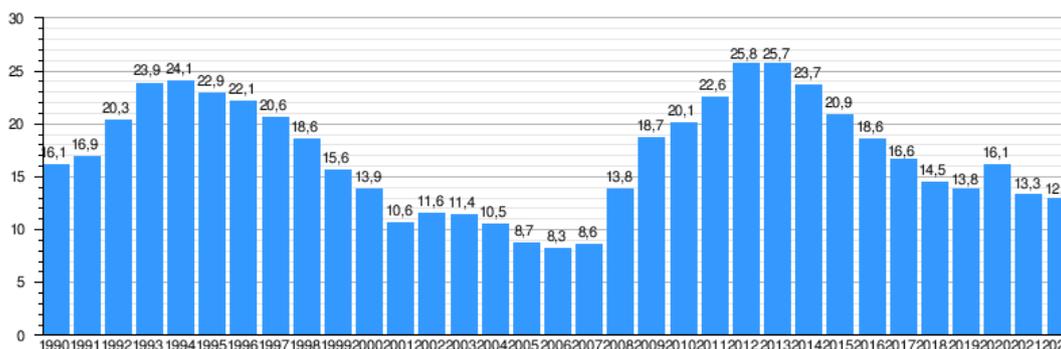


Gráfico 1 Evolución anual de la tasa de paro en España. Fuente: INE

Aun así, cabe destacar que esta caída en la tasa de paro no es un caso aislado de España, si no que está muy relacionado con el desarrollo económico europeo, por lo que la diferencia en el porcentaje de personas desempleadas en España con relación a la UE y la eurozona es bastante similar a la de otros años.

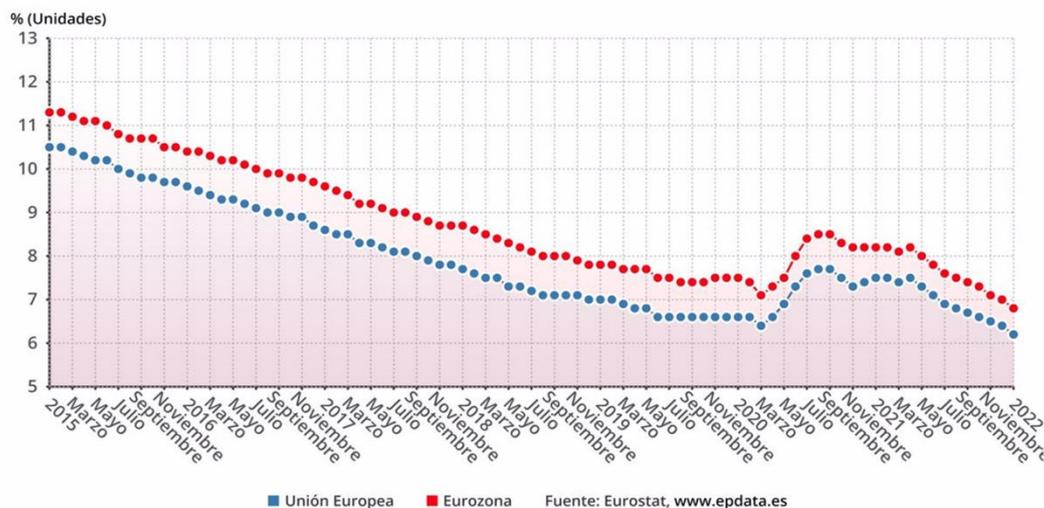
¹ «Barómetro del CIS marzo 2022» Centro de Investigaciones sociológicas.

² «El paro de la eurozona continúa en mínimos históricos con un 6,5%. España continúa a la cabeza de la lista, con una tasa del 12,4%». Cinco Días. 9 de enero de 2023.

³ Datos del último trimestre de 2022 obtenidos del Instituto Nacional de Estadística

Evolución de la tasa de desempleo en la UE y en la Eurozona

Datos desde enero de 2015



4

Gráfico 2 Evolución de la tasa de desempleo en la UE y en la Eurozona.

Algunas de las causas más importantes del desempleo estructural en España son:

- Terciarización de la economía⁵.

Este es un proceso de cambio en la estructura económica española, que se ha transformado de una economía basada en los sectores de actividad primaria, como eran la agricultura y la ganadería, a una nueva organización en la que son los sectores de servicios los que por su magnitud toman mayor importancia.

Tanto es así que el sector terciario suponía en el año 2020 un 68,5% del PIB nacional, frente al 14,6% y 2,2% de los sectores industrial y agrícola respectivamente.

- Rigideces en el mercado laboral y alto grado de temporalidad.

Estas rigideces se traducen en la dificultad para la creación de empleo en nuestro país, por la complejidad de las empresas de adaptarse a la variación en la demanda y la tecnología. Algunas de estas rigideces son

⁴ Eurostat: "Total unemployment rate".

⁵ Cuadrado Roura, Juan Ramón «La terciarización de la economía española: algunos rasgos estilizados del período 1975-2014».

la alta temporalidad, los costes de despido y la rigidez salarial a consecuencia de los convenios.

En 1975 la tasa de paro en España era inferior a la de Estados Unidos y estaba a la par con la de la Unión Europea⁶. Sin embargo, era un falso equilibrio ya que la mayoría de las mujeres no se habían incorporado al mercado laboral, y muchos hombres emigraron para trabajar en el extranjero. Por lo tanto, la oferta de trabajo era mucho menor.

En el periodo 1975-1985, se produjo una fuerte destrucción de empleo que hizo que aumentara la tasa de paro de manera exponencial. La oferta de trabajo también aumentó, por lo que contribuyó al aumento del paro. La evolución de la oferta de trabajo en este periodo fue resultado de dos tendencias contrapuestas:

- El aumento de la población potencialmente activa debido a las altas tasas de natalidad en las décadas anteriores y al retorno de los emigrantes.
- Caída de la tasa de actividad debido a que las expectativas de encontrar empleo eran cada vez menores y una parte importante de la población mayor de 16 años dejó de buscar trabajo.

En la etapa 1986-1991, aumenta la demanda de trabajo y el paro se reduce, sin embargo, no es capaz de volver al origen, ya que también aumenta la oferta de trabajo. La oferta de trabajo creció debido a la llegada del baby boom al mercado de trabajo, a que hubo población que volvió a buscar trabajo gracias a la mejora de las expectativas para encontrarlo y a la incorporación masiva de la mujer en el mercado de trabajo.

En el periodo 1992-1994 se destruye la mitad del empleo que se había creado en la etapa anterior. Además, como la oferta sigue creciendo, la tasa de paro se volvió a disparar.

En el periodo 1995-2007, la demanda de trabajo aumentó en gran medida, se crearon muchos puestos de trabajo. La oferta también aumenta gracias a la

⁶ Sanchez-Silvia, Carmen «Historia de un paro que no cesa». El País.

llegada de inmigrantes, pero no tanto como la demanda, por lo que la tasa de paro se reduce.

En la etapa 2008-2013, se destruyen muchos puestos de trabajo, gran parte correspondientes al sector de la construcción. La oferta de trabajo sigue aumentando porque la tasa de actividad rompe su tendencia procíclica, ya que la población no se podía permitir dejar de buscar trabajo debido a la fuerte crisis económica.

Desde 2014 hasta la actualidad, se vuelve a crear empleo y la oferta de trabajo disminuye débilmente, ya que muchos inmigrantes volvieron y otros dejan de buscar trabajo ante las mejores expectativas.

Por último, cabe destacar la influencia del Covid-19 en un aumento puntual de la tasa de paro en el año 2020. Sin embargo, correspondió únicamente a un momento concreto y no afectó al desempleo estructural de la economía, pues los años 2021 y 2022 se ha continuado con la buena tendencia de descenso del desempleo.

La Encuesta de Población Activa (EPA) es una investigación que se realiza de manera trimestral con el fin de definir y estudiar cuantitativamente el mercado de trabajo en España en base a diferentes factores.

Todos los países de la Unión Europea están obligados a la realización de la encuesta de población activa. Para el caso de España, es el Instituto Nacional de Estadística (INE) el que se encarga de esta actividad.

Esta encuesta recoge un total de noventa variables, de un número aproximado de 160.000 individuos de unos 65.000 hogares y puede ser realizada personalmente o por vía telefónica.

Acorde a la Ley 12/1989, de 9 de mayo, de la Función Estadística Pública, los servicios estadísticos podrán solicitar datos de todas las personas físicas y jurídicas, nacionales y extranjeras, residentes en España.

De la misma manera, recogido en la misma ley se obliga a todas las personas físicas y jurídicas que suministren datos, tanto si su colaboración es obligatoria como voluntaria, a contestar de forma veraz, exacta, completa y dentro del plazo

a las preguntas ordenadas en la debida forma por parte de los servicios estadísticos.

Para la realización de este trabajo, los datos utilizados son correspondientes a la Encuesta de Población Activa (EPA) del primer trimestre de 2022.

El objetivo de este trabajo es estudiar como ciertas variables personales influyen en la probabilidad de estar desempleado en España. Estas variables son la edad, el sexo, el nivel de estudios, el estado civil y la nacionalidad

Se analizará si estas variables influyen y en qué medida a la hora de conseguir un empleo en nuestro país. También se estudiará posibles interacciones entre las variables, a modo de poder afirmar que existan diferencias para individuos pertenecientes a segmentos específicos.

2. ANÁLISIS DESCRIPTIVO DE LAS VARIABLES

Antes de la aplicación del modelo Logit, se va a realizar un análisis descriptivo de cada variable empleada, con objetivo de definir cuantitativamente el número de personas desempleadas pertenecientes a cada grupo, además de diversas tablas cruzadas entre variables, para posteriormente estudiar las diferencias encontradas para diferentes perfiles de individuos.

Los datos para este estudio han sido obtenidos del Instituto Nacional de Estadística (INE) y corresponden a la Encuesta de Población Activa (EPA) del primer trimestre de 2022.

Como se ha mencionado anteriormente, la Encuesta de Población Activa (EPA), utiliza un total de 90 variables para el estudio y análisis del desempleo en España. Sin embargo, para la realización de este estudio tomaremos en cuenta únicamente las variables sexo, nivel de estudios, estado civil, edad, nacionalidad y situación laboral.

Además, se han eliminado de la muestra a todos aquellos individuos cuya situación laboral es inactiva, ya que no son necesarios para el estudio.

Para la muestra final que vamos a utilizar, nos hemos quedado con un total de 60.370 individuos.

2.1. Metodología

Para el estudio descriptivo de las variables utilizaremos análisis de frecuencias y tablas cruzadas a través de la herramienta de estudio estadístico SPSS, con el objetivo de cuantificar el número y porcentaje de ocupados y desocupados dentro de las categorías de cada variable.

Además, compararemos diferentes variables entre sí para ver las diferencias entre distintas alternativas en función de otra variable, de manera que incluimos otra variable al estudio para analizar las diferencias en el porcentaje de desempleo para perfiles más concretos. Por ejemplo, utilizaremos las variables sexo y estado civil, para ver el porcentaje de desempleo en hombres solteros y casados y mujeres solteras y casadas por separado

2.2. Variables

2.2.1. Situación laboral

La situación laboral de los encuestados está definida por la variable **AOI** y será la variable dependiente del posterior estudio mediante el modelo Logit.

Esta variable está codificada de la siguiente manera:

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

Es importante distinguir diferentes conceptos a la hora de referirse a la situación laboral de los individuos:

- Población potencialmente activa: población de 16 años o más.
- Población ocupada: mayores de 16 años que realizan una actividad económica remunerada \approx demanda de trabajo.

Puede darse la situación en la que la demanda de trabajo sea mayor que la población ocupada, ya que existe la posibilidad de que ciertos sectores tengan salarios más bajos para los que los individuos no estén dispuestos a trabajar y no se cubra la demanda. El ejemplo más común es la hostelería.

- Población parada: mayores de 16 años que están sin trabajo, lo buscan activamente y están disponibles para trabajar.
- Población activa: población ocupada + población parada = oferta de trabajo.
- Población inactiva: mayores de 16 años que no desarrollan una actividad económica remunerada, bien porque no están disponibles para trabajar, o bien porque, estando sin trabajo, no lo buscan activamente.

Cuando nos referimos a la tasa de paro, hacemos referencia al cociente entre la población parada y la población activa, pues corresponde al porcentaje de personas dispuestas a trabajar que no disponen de un empleo.

Para nuestro análisis solo nos interesan las personas activas, es decir, mayores de 16 años que son aptas para trabajar y están dispuestos a ello. En consecuencia, las variables 07, 08 y 09 que corresponden a los individuos entrevistados inactivos son eliminadas de la muestra.

Por otro lado, para este estudio solo necesitamos distinguir entre población ocupada y desempleada, por lo que se ha recodificado la variable en una nueva, **AOI1**, de la siguiente manera:

- 0: Ocupados (individuos pertenecientes al grupo 03 y 04)
- 1: Desocupados (individuos pertenecientes al grupo 05 y 06)

		AOI1			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Ocupado	52502	87,0	87,0	87,0
	Desocupado	7868	13,0	13,0	100,0
	Total	60370	100,0	100,0	

Tabla 1 Tabla de frecuencias de la variable Situación laboral.

Por medio del análisis de frecuencias de la *Tabla 1* observamos que, de los 60.370 individuos de la muestra, EL 13% se encuentran en situación de desempleo, frente al 87% ocupados.

La variable **AOI1** será la que utilizaremos para el resto del análisis para definir la situación laboral de los individuos.

2.2.2. Edad

La variable **EDAD1** recoge la edad en tramos de 5 años (a excepción del tramo 16-19) de los individuos entre 16 y 64 años. A partir de dicha edad considera a todos los individuos de la muestra en un único segmento (mayores de 65).

La variable está codificada de la siguiente manera:

- 16: Individuos entre 16 y 19 años
- 20: Individuos entre 20 y 24 años
- 25: Individuos entre 25 y 29 años
- 30: Individuos entre 30 y 34 años
- 35: Individuos entre 35 y 39 años
- 40: Individuos entre 40 y 44 años
- 45: Individuos entre 45 y 49 años
- 50: Individuos entre 50 y 54 años
- 55: Individuos entre 55 y 59 años
- 60: Individuos entre 60 y 64 años
- 65: Individuos de 65 años o más

Para el análisis descriptivo, hemos creado una nueva variable a partir de **EDAD1**, llamada **EDAD2**, que agrupa los individuos en tramos de edad más grandes para simplificar la muestra. Los nuevos segmentos de edad son:

- 16 a 24 años
- 25 a 44 años
- 45 a 59 años
- 60 o más años

Tras la realización de una tabla cruzada entre las variables **EDAD2** y **AOI1**, nos percatamos de la siguiente distribución del desempleo por cada grupo edad de los individuos:

		PORCENTAJE DE DESEMPLEO	RECuento DE DESEMPLEADOS	TOTAL DE INDIVIDUOS
EDAD	16 a 24 años	31,2%	1144	3663
	25 a 44 años	13,1%	3106	23672
	45 a 59 años	11,0%	2950	26927
	60 o más años	10,9%	668	6108

Tabla 2 Tabla cruzada entre las variables de Edad y Situación laboral. Fuente: elaboración propia a partir de los datos del INE

Rápidamente podemos observar en la *Tabla 2* que el porcentaje de desempleo es muy superior entre los menores de 25 años (con un 31,2% de desocupados) que del resto de tramos, que se encuentran mucho más cerca de la media total de la muestra del 13%.

También podemos ver que el porcentaje de desempleo es menor a medida que nos acercamos a grupos de mayor edad, siendo la tasa de paro más baja la que corresponde a los mayores de 60 años.

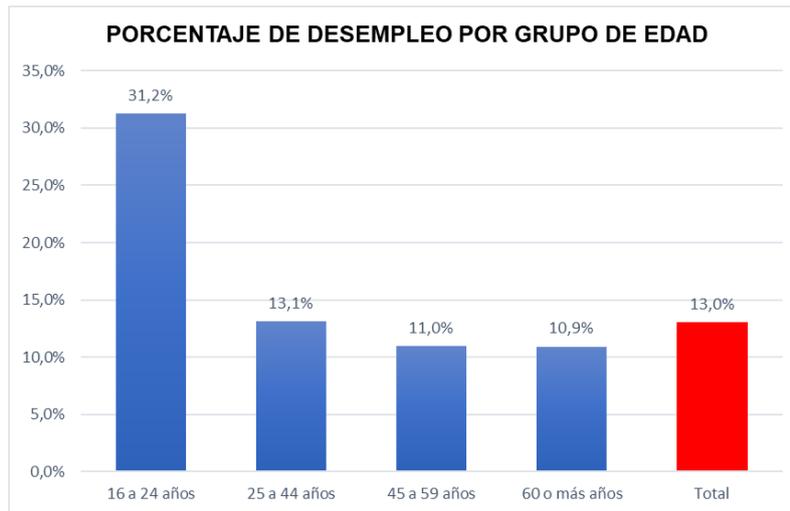


Gráfico 3 Porcentaje de desempleo por grupo de edad de la muestra. Fuente: elaboración propia a partir de los datos del INE

A partir de la variable **EDAD1**, también se ha creado una nueva variable continua, **EDADCONT**, que servirá para poder considerar la edad de los individuos de manera numérica en lugar de por tramos de edad, lo que no será perfectamente realista debido a que no refleja la edad real de cada individuo, si no la media del grupo al que pertenece, pero es lo más aproximado a la realidad que podemos obtener utilizando los datos que tenemos.

Al ser una variable continua, no existe codificación alguna para la misma, directamente refleja la edad de los individuos.

2.2.3. Sexo

La variable **SEXO1** corresponde al género de los individuos y viene codificada de la siguiente manera:

- 1: Varón
- 0: Mujer

		PORCENTAJE DE DESEMPLEO	RECUESTO DE DESEMPLEADOS	TOTAL DE INDIVIDUOS
SEXO	Hombre	11,3%	3523	31237
	Mujer	14,9%	4345	29133

Tabla 3 Tabla cruzada entre las variables Sexo y Situación laboral. Fuente: elaboración propia a partir de los datos de INE

El recuento de individuos realizado por medio de una tabla cruzada en SPSS nos indica que el 51,7% de la muestra utilizada corresponde a varones y el 48,3% a mujeres, reflejado así en la *Tabla 3*.

Sin embargo, el recuento refleja una mayor cantidad de mujeres paradas que de hombres a pesar de ser un menor número de encuestadas. En consecuencia, también podemos observar rápidamente en el *Gráfico 4* que el porcentaje de personas desocupadas es mayor en el género femenino (14,9% en mujeres y 11,3% en hombres).

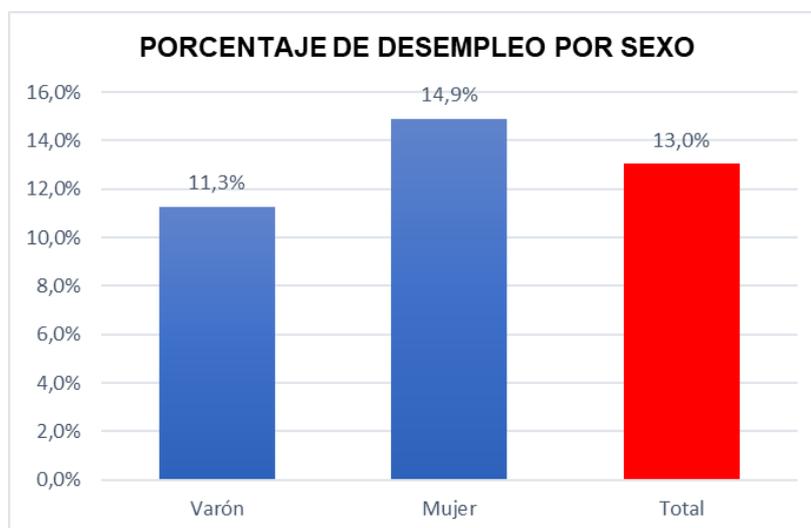


Gráfico 4 Porcentaje de desempleo por sexo de la muestra. Fuente: elaboración propia a partir de los datos del INE

2.2.4. Nivel de estudios

La variable **NFORMA** hace referencia al nivel de estudios, con 7 alternativas posibles. Sin embargo, hemos recodificado esta variable en una nueva (**NFORMA1**) de la siguiente manera:

- 0: Sin estudios → Hemos incluido analfabetos (AN) y educación primaria no terminada (P1)
- 1: Educación primaria → P2
- 2: Educación secundaria → S1
- 3: Bachillerato o equivalente → Unido las dos variables de segunda etapa de educación secundaria (SG Y SP).
- 4: Estudios superiores → SU

		PORCENTAJE DE DESEMPLEO	RECuento DE DESEMPLEADOS	TOTAL DE INDIVIDUOS
NIVEL DE ESTUDIOS	Sin estudios	30,6%	187	611
	Educación primaria	28,1%	605	2151
	Educación secundaria	17,7%	3014	17053
	Bachillerato o equivalente	14,3%	2059	14415
	Estudios superiores	7,7%	2003	26140

Tabla 4 Tabla cruzada entre las variables Nivel de estudios y Situación laboral. Fuente: elaboración propia a partir de los datos del INE

Como podemos observar en la *Tabla 4*, la gran mayoría de los encuestados poseen estudios de educación secundaria obligatoria o superior y el porcentaje de personas en situación de desempleo es menor a mayor nivel de estudios.

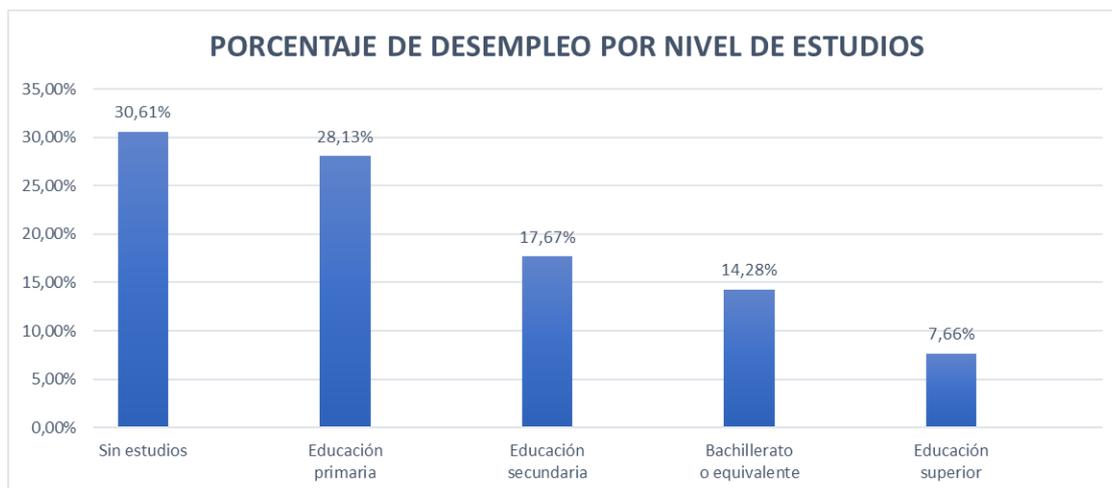


Gráfico 5 Porcentaje de desempleo por nivel de estudios de la muestra. Fuente: elaboración propia a partir de los datos del INE

Solamente el grupo de individuos con estudios superiores tiene una tasa de paro por debajo de la media de la muestra (13%).

Si introducimos la variable **SEXO1** en el análisis, podemos ver las diferencias en el porcentaje de desempleo para cada nivel de estudios en hombres y mujeres.

		PORCENTAJE DE DESEMPLEO	RECuento DE DESEMPLEADOS	TOTAL DE INDIVIDUOS
Hombre	Sin estudios	26,5%	102	385
	Educación primaria	25,1%	330	1316
	Educación secundaria	14,2%	1455	10230
	Bachillerato o equivalente	11,9%	899	7530
	Estudios superiores	6,3%	737	11776
Mujer	Sin estudios	37,6%	85	226
	Educación primaria	32,9%	275	835
	Educación secundaria	22,8%	1559	6823
	Bachillerato o equivalente	16,8%	1160	6885
	Estudios superiores	8,8%	1266	14364

Tabla 5 Tabla cruzada entre las variables Nivel de estudios, Sexo y Situación laboral. Fuente: Elaboración propia a partir de los datos del INE

Acorde a la *Tabla 5*, podemos observar que las mujeres tienen mayor tasa de paro para todos los niveles de estudios. Por otro lado, se refleja una mayor

cantidad de mujeres con estudios superiores que los varones, mientras que son los hombres los que tienen mayor presencia en el resto de las categorías.

En ambos casos, los individuos con estudios superiores tienen una probabilidad por debajo de la media de la muestra de encontrarse en situación de desempleo.

2.2.5. Nacionalidad

La variable **NAC1** refleja la nacionalidad de los individuos, distinguiendo entre:

- 1: Español
- 2: Español con doble nacionalidad
- 3: Extranjero

Para el estudio, hemos recodificado la variable para tener únicamente dos alternativas, considerando a los españoles con doble nacionalidad como españoles. La nueva variable, **NAC2**, queda de la siguiente manera:

- 0: Extranjeros
- 1: Españoles

		PORCENTAJE DE DESEMPLEO	RECuento DE DESEMPLEADOS	TOTAL DE INDIVIDUOS
NACIONALIDAD	Español	12,3%	6886	55918
	Extranjero	22,1%	982	4452

Tabla 6 Tabla cruzada de las variables Nacionalidad y Situación laboral. Fuente: elaboración propia a partir de los datos del INE

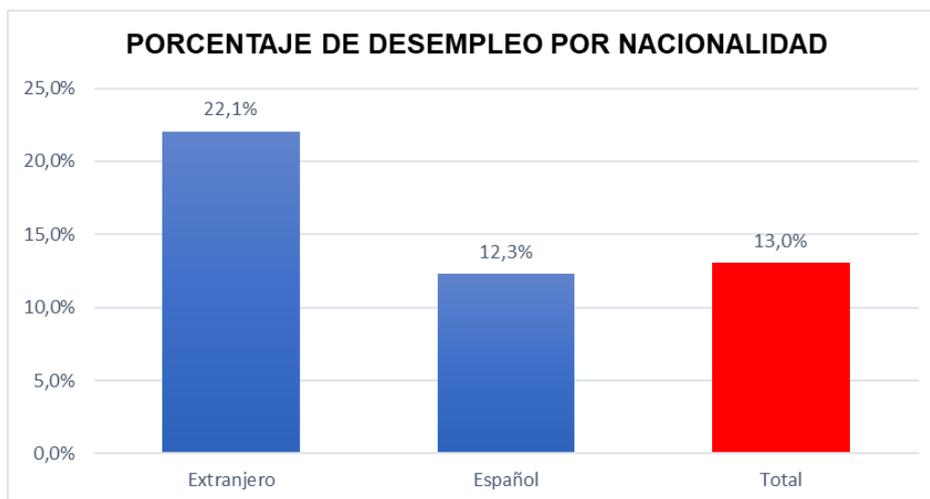


Gráfico 6 Porcentaje de desempleo por nacionalidad de la muestra. Fuente: elaboración propia a partir de los datos del INE

Como podemos observar en la *Tabla 6* y *Gráfico 6*, la mayor parte de los individuos son de nacionalidad española, entre los cuales, la tasa de paro es inferior (12,3%) con respecto a los extranjeros (22,1%).

A continuación utilizaremos el sexo para distinguir el porcentaje de desempleados entre hombres españoles y extranjeros y mujeres españolas y extranjeras.

		PORCENTAJE DE DESEMPLEO	RECuento DE DESEMPLEADOS	TOTAL DE INDIVIDUOS
Hombre	Español	10,7%	3097	28962
	Extranjero	18,7%	426	2275
Mujer	Español	14,1%	3789	26956
	Extranjero	25,5%	556	2177

Tabla 7 Tabla cruzada de las variables Nacionalidad, Sexo y Situación laboral. Fuente: elaboración propia a partir de los datos del INE

Introduciendo la variable **SEXO1**, observamos en la *Tabla 7* que los resultados apoyan los ya obtenidos al estudiar el desempleo de manera aislada las variables sexo y nacionalidad, ya que la tendencia sigue desfavoreciendo a mujeres y extranjeros.

2.2.6. Estado civil

El estado civil de los individuos se codifica originalmente en la variable **ECIV1**, distinguiendo entre solteros, casados, viudos y divorciados. Para los intereses de nuestro análisis solo necesitamos distinguir entre los individuos solteros y casados, por lo que reconoceremos a viudos y divorciados como solteros.

Para ello, hemos recodificado **ECIV1** en una nueva variable, **ECIV2**:

- 0: Casado
- 1: Soltero

		PORCENTAJE DE DESEMPLEO	RECUENTO DE DESEMPLEADOS	TOTAL DE INDIVIDUOS
ESTADO CIVIL	Casado	9,4%	3044	32349
	Soltero	17,2%	4824	28021

Tabla 8 Tabla cruzada entre las variables Estado civil y Situación laboral. Fuente: elaboración propia a partir de los datos del INE

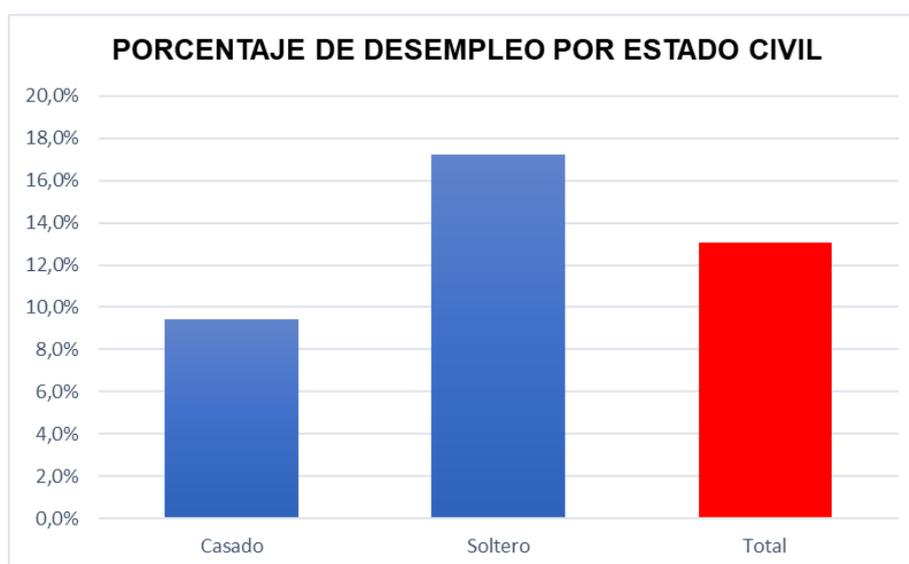


Gráfico 7 Porcentaje de desempleo por estado civil de la muestra. Fuente: elaboración propia a partir de los datos del INE

Los resultados descriptivos reflejados en el *Gráfico 7* indican que la tasa de paro es considerablemente mayor entre personas solteras (17,2%) que en las personas casadas (9,4%).

A continuación, utilizaremos la variable sexo para ver las diferencias en la tasa de paro entre hombre solteros y casados y mujeres solteras y casadas:

		PORCENTAJE DE DESEMPLEO	RECuento DE DESEMPLEADOS	TOTAL DE INDIVIDUOS
Hombre	Casado	6,5%	1111	17036
	Soltero	17,0%	2412	14201
Mujer	Casado	12,6%	1933	15313
	Soltero	17,5%	2412	13820

Tabla 9 Tabla cruzada entre las variables Estado civil, Sexo y Situación laboral. Fuente: elaboración propia a partir de los datos del INE

Por otro lado, tras la realización de la *Tabla 9*, incluyendo la variable **SEXO1**, podemos observar que la tasa de paro es considerablemente mayor entre mujeres casadas (12,6%) que entre hombres casados (6,5%), mientras que, entre los solteros, la tasa de paro es muy similar en ambos sexos.

3. MODELO LOGIT

3.1. Introducción teórica

El modelo Logit es un modelo estadístico de elección discreta, en el que la variable dependiente es dicotómica, solo toma dos valores. Los modelos de elección discreta explican una variable cualitativa en función de otras variables, bien sean cualitativas o cuantitativas.

Para el estudio que pretendemos realizar, es más apropiado el uso de este modelo por encima del modelo lineal de probabilidad, ya que este último puede darnos valores que no se encuentren comprendidos entre 0 y 1.

El modelo Logit soluciona este problema transformando el modelo lineal de probabilidad en una función comprendida entre esos valores de la siguiente manera:

$$Y_i = F(\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}) + u_i^7$$

Dónde $0 \leq F \leq 1$ e Y_i es una variable dicotómica que recoge si el individuo está desempleado o no y cuya media representa la probabilidad de que un individuo se encuentre en situación de desempleo en función del valor resto de variables.

En nuestro estudio, el valor 0 corresponde a la situación de ocupado y el valor 1 al desempleo.

Expresado en términos matemáticos, la esperanza de Y_i es igual a la probabilidad de que Y sea igual a 1, que a su vez es igual a la función $F(X_i\beta)$:

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

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

F es una función de distribución logística, con esperanza igual a cero y la desviación típica es $\frac{\pi}{\sqrt{3}}$. La función por tanto es la siguiente:

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

$$E(Y_i) = p(Y_i = 1) = p_i = \frac{e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}}}{1 + e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}}}$$

$$p(Y_i = 0) = 1 - p_i = \frac{1}{1 + e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}}}$$

$$\ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}$$

En el modelo Logit, a diferencia del modelo de regresión lineal, no medimos cómo cambia la probabilidad de que $Y_i = 1$ al aumentar en una unidad cada variable independiente, ya que no es un modelo lineal, por lo que la probabilidad de $Y_i = 1$ y el efecto marginal no depende únicamente del valor de las β , sino también de las X, por lo que no es constante.

⁷ Prieto Alaiz, Mercedes. Formulas obtenidas de los apuntes de clase de la asignatura de Análisis Multivariante de Datos Económicos del curso 2021-2022.

$$\Delta p_i = F(\beta_0 + \beta_1 X_{1i} + \dots + \beta_j (X_{ji} + 1) + \dots + \beta_k X_{ki}) - F(\beta_0 + \beta_1 X_{1i} + \dots + \beta_j X_{ji} + \dots + \beta_k X_{ki}) \neq \beta_j$$

Por otro lado, tenemos que el efecto marginal del regresor X_j no es constante:

$$\frac{\partial p_i}{\partial X_{1i}} = \frac{\partial F(t)}{\partial t} \frac{\partial t}{\partial X_{1i}} = f(\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}) \beta_j = f(X_i' \beta) \beta_j \neq \beta_j$$

Por tanto, dado que $\partial p / \partial X_{ji} = f(X_i' \beta) \beta_j$, en el modelo logit el efecto marginal tiene la siguiente expresión:

$$\frac{\partial p_i}{\partial X_{ji}} = F(X_i' \beta) [1 - F(X_i' \beta)] \beta_j = p_i (1 - p_i) \beta_j$$

Otro concepto que debemos tener en cuenta es el *Odd*, que hace referencia al cociente entre la probabilidad de $Y_i = 1$ e $Y_i = 0$. Para valores superiores a 1, la probabilidad de encontrarse en situación de desempleo será mayor a la de estar ocupado.

$$O_i = \frac{p(Y_i = 1)}{p(Y_i = 0)} = \frac{\frac{e^{X_i' \beta}}{1 + e^{X_i' \beta}}}{\frac{1}{1 + e^{X_i' \beta}}} = e^{X_i' \beta} = e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}}$$

Por último, tenemos el Odd Ratio, que es el cociente entre el Odd al variar una unidad de una de las variables manteniendo igual al resto con respecto al Odd original.

$$\frac{O(\beta_0 + \beta_1 X_{1i} + \dots + \beta_j (X_{ji} + 1) + \dots + \beta_k X_{ki})}{O(\beta_0 + \beta_1 X_{1i} + \dots + \beta_j X_{ji} + \dots + \beta_k X_{ki})} = e^{\beta_j}$$

3.2. Aplicación del modelo Logit

Para la aplicación del modelo Logit se ha necesitado crear variables ficticias para la variable Nivel de estudios. Estas nuevas variables son, **SECUNDARIA**, **BACH** y **SUPERIOR**, que toman valor 1 si el individuo posee ese nivel de estudios y el valor 0 si posee cualquier otro nivel de estudios. Por ejemplo, para la variable **SECUNDARIA**, los individuos que posean educación secundaria tendrán

asignado el valor 1, mientras que todos aquellos que bien no posean estudios primarios o que hayan alcanzado un nivel superior como son secundaria, bachillerato o equivalente o estudios superiores tendrán asignado el valor 0.

Los individuos sin estudios o con estudios primarios tomarán el valor 0 en todas las variables. Han sido agrupadas ya que son muy pocos los individuos que pertenecen a estos grupos

Para una determinada variable con n categorías, se pueden crear como máximo $n-1$ variables ficticias, por lo que la manera de recoger a todos aquellos individuos que no poseen ningún tipo de nivel de estudios será tomando el valor 0 en todas las variables ficticias relativas al nivel de educación.

El análisis Logit lo hemos realizado en SPSS a través de la siguiente ruta:

Analizar → Regresión → Logística binaria

En la ventana de selección de variables hemos introducido la siguiente combinación:

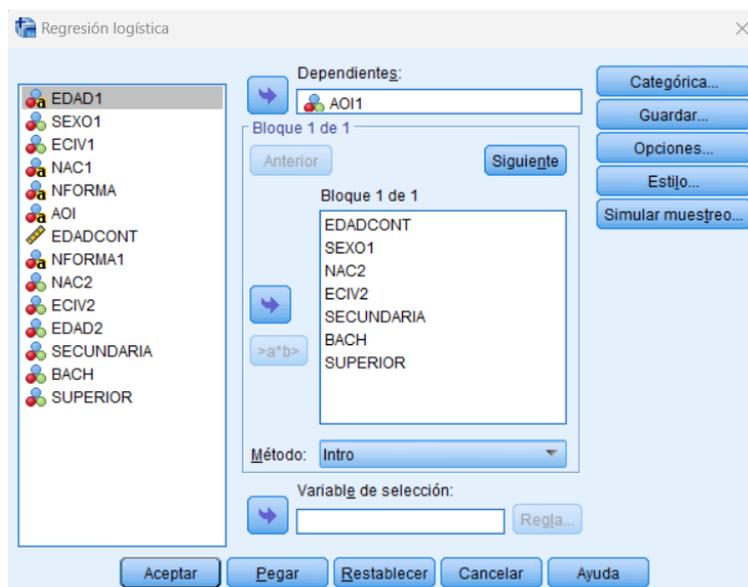


Ilustración 2 Ventana de selección de variables en el análisis Logit en SPSS.

Los resultados obtenidos en el análisis son los siguientes:

Variables en la ecuación

		B	Error estándar	Wald	gl	Sig.	Exp(B)
Paso 1 ^a	EDADCONT	-,022	,001	378,392	1	,000	,978
	SEX01	-,480	,025	357,678	1	,000	,619
	NAC2	-,360	,041	76,993	1	,000	,698
	ECIV2	,525	,028	354,447	1	,000	1,690
	SECUNDARIA	-,640	,049	171,981	1	,000	,527
	BACH	-1,032	,051	413,716	1	,000	,356
	SUPERIOR	-1,731	,051	1148,896	1	,000	,177
	Constante	,428	,078	30,183	1	,000	1,534

a. Variables especificadas en el paso 1: EDADCONT, SEX01, NAC2, ECIV2, SECUNDARIA, BACH, SUPERIOR.

Tabla 10 Tabla de resultados del análisis Logit en SPSS.

Con los valores de la columna Exp(B) podemos ver el cociente entre los Odd de las dos categorías de cada variable manteniendo el resto constante, es decir, esta columna nos indica el Odd Ratio de cada variable. Por ejemplo, en **SUPERIOR**, el valor de 0,177 nos indica que poseer este tipo de estudios reduce la razón entre la probabilidad de estar desempleado y no estarlo se reduce un 82,3% con respecto a los individuos sin estudios o con estudios primarios.

Para el caso de la variable **EDADCONT**, el valor del Odd Ratio es muy similar a 1, por lo que apenas existen diferencias al variar en una unidad la edad de los individuos permaneciendo el resto de las variables constantes.

El estado civil (**ECIV2**) también implica grandes diferencias, aumentando un 69% la razón entre la probabilidad de estar desempleado y no estarlo en individuos solteros con respecto a los casados manteniendo el resto de las variables constantes.

El valor de la columna *B* nos indica el coeficiente de cada variable en el modelo Logit, sin embargo, no debemos tomar este valor como la cantidad en que incrementa la probabilidad de estar en situación de desempleo en función de cada variable, y solo tomaremos en cuenta el signo de este coeficiente para ver si las variables tienen una relación directa o inversa con la dependiente del modelo.

En primer lugar, observamos que la variable **EDADCONT** tiene coeficiente negativo, de manera que, a medida que aumenta la edad de los individuos menor será la probabilidad de encontrarse en situación de desempleo.

Para el resto de las variables dicotómicas podemos sacar el mismo tipo de conclusiones:

- **SEXO1**: Es más probable encontrarse en situación de desempleo si eres mujer con respecto a los hombres.
- **NAC2**: Es más probable encontrarse en situación de desempleo si eres extranjero con respecto a los individuos de nacionalidad española.
- **ECIV2**: Es más probable encontrarse en situación de desempleo si eres soltero con respecto a los casados.
- Individuos con estudios secundarios, bachillerato o equivalente o estudios superiores tienen una menor probabilidad de encontrarse en situación de desempleo que los individuos sin estudios o con estudios primarios.

Una vez vista la relación de las variables independientes con la variable dependiente hay que valorar su significación, para determinar si son o no influyentes para el análisis.

Como podemos observar en el p-valor, todas las variables son significativas para un p-valor de 0,05.

También hemos obtenido la siguiente tabla de clasificación de pronóstico de los casos en función de las variables:

Tabla de clasificación^a

Observado		Pronosticado		Porcentaje correcto	
		Ocupado	AOI1 Desocupado		
Paso 1	AOI1	Ocupado	52437	65	99,9
		Desocupado	7818	50	,6
		Porcentaje global			86,9

a. El valor de corte es ,500

Tabla 11 Tabla de clasificación de los casos acorde al modelo Logit.

Como podemos observar en la *Tabla 11* SPSS ha pronosticado como parados a todos aquellos con un $p(Y_i = 1) > 0,5$ logrando un porcentaje global de aciertos del 86,9%.

Sin embargo, únicamente se ha pronosticado de manera acertada un 0,6% de los individuos desempleados y lo que el gran porcentaje de acierto se debe a que ha clasificado a casi todos los individuos como ocupados, lo que corresponde a un porcentaje muy alto de la muestra utilizada.

Para conseguir una mejor clasificación, hemos reducido el valor de corte al 13%, correspondiente al porcentaje de parados de la muestra que estamos utilizando para que clasifique como parados a aquellos con una $p(Y_i = 1) > 0,13$, de manera que esperamos obtener una mejor clasificación.

Tabla de clasificación^a

Observado		Pronosticado		Porcentaje correcto	
		Ocupado	AOI1 Desocupado		
Paso 1	AOI1	Ocupado	33346	19156	63,5
		Desocupado	2653	5215	66,3
		Porcentaje global			63,9

a. El valor de corte es ,130

Tabla 12 Tabla de clasificación de los casos acorde al modelo Logit para $p(Y_i = 1) > 0,13$.

En esta nueva clasificación, el porcentaje de acierto global es menor, del 63,9%, sin embargo, el porcentaje de acierto en individuos ocupados y desocupados es bastante más similar.

4. ESCENARIOS

A continuación, aplicaremos la fórmula del modelo Logit en casos concretos, utilizando los coeficientes obtenidos y valores específicos para las X, con el fin de obtener la probabilidad de estar desempleado en diferentes escenarios y compararlos entre sí.

EDAD	SEXO	NACIONALIDAD	ESTADO CIVIL	NIVEL DE ESTUDIOS	PROBABILIDAD DE DESEMPLEO
20	Hombre	Español	Soltero	Secundaria	0,2755
20	Hombre	Extranjero	Soltero	Secundaria	0,3527
20	Mujer	Español	Soltero	Secundaria	0,3806
20	Mujer	Extranjero	Soltero	Secundaria	0,4683
25	Hombre	Español	Soltero	Superiores	0,1027
25	Hombre	Español	Soltero	Bachillerato	0,1871
25	Mujer	Español	Soltero	Superiores	0,1560
25	Mujer	Español	Soltero	Bachillerato	0,2711
30	Hombre	Español	Soltero	Primaria	0,3666
30	Hombre	Extranjero	Soltero	Primaria	0,4534
30	Hombre	Español	Soltero	Superiores	0,0930
30	Hombre	Extranjero	Soltero	Superiores	0,1281
35	Hombre	Español	Soltero	Secundaria	0,2147
35	Hombre	Español	Casado	Secundaria	0,1392
35	Mujer	Español	Casado	Secundaria	0,2072
40	Mujer	Español	Soltero	Secundaria	0,2835
40	Mujer	Español	Casado	Secundaria	0,1897
40	Mujer	Español	Casado	Superiores	0,0729
40	Mujer	Extranjero	Soltero	Primaria	0,5182
40	Mujer	Extranjero	Soltero	Superiores	0,1600
40	Mujer	Extranjero	Casado	Primaria	0,3889
50	Hombre	Español	Casado	Secundaria	0,1041
50	Mujer	Español	Casado	Secundaria	0,1582
50	Mujer	Español	Soltero	Secundaria	0,2410
60	Hombre	Español	Casado	Primaria	0,1503
60	Mujer	Español	Casado	Primaria	0,2224
60	Hombre	Español	Casado	Superiores	0,0304

Tabla 13 Tabla de escenarios. Fuente: elaboración propia

Hemos considerado un total de 27 escenarios, con algunos casos particulares que podrían ser influyentes en la probabilidad de encontrarse en situación de desempleo.

El perfil de individuo con mayor probabilidad de encontrarse en situación de desempleo de los reflejados en los escenarios corresponde a aquellos varones españoles de 60 años, casados y con estudios superiores, con tan solo un 0,0304 de probabilidad.

El perfil de individuo con menor probabilidad de encontrarse en situación de desempleo de los reflejados en los escenarios son mujeres extranjeras de 40 años, solteras y con estudios primarios, con un 0,5182 de probabilidad.

En cuanto a las variables, podemos observar que hay dos que influyen de manera contundente en el desempleo: la nacionalidad extranjera y los estudios superiores. Los casos en los que los individuos tienen nacionalidad extranjera reflejan una probabilidad más alta de no tener empleo en comparación con otro individuo con el resto de las variables iguales. Por el contrario, los estudios superiores tienen gran influencia en la probabilidad de tener trabajo, reflejando probabilidades mucho más bajas los casos en los que los individuos tienen este tipo de estudios en comparación a casos con el resto de las variables idénticas pero menor nivel de estudios.

En personas de 20 años, solteras y con educación secundaria; la probabilidad de sufrir desempleo es mayor en mujeres, independientemente de la nacionalidad, existiendo una probabilidad estimada de 0,4683 de ser desempleada para mujeres extranjeras de este perfil.

Un caso muy interesante resulta de valorar la probabilidad de no encontrar trabajo al terminar la carrera, máster o estudios equivalentes. Para los casos más comunes: jóvenes españoles de unos 25 años, solteros y como hemos dicho, con estudios superiores, podemos observar que la probabilidad de encontrarse en situación de desempleo es menor que la de aquellos que han estudiado bachillerato, con un 0,1027 en hombres y un 0,1560 en mujeres, frente al 0,1871 y 0,2711 respectivos.

También podemos comparar individuos con las mismas características a excepción del estado civil. En todos los escenarios que hemos estimado, los solteros presentan una mayor probabilidad de no tener empleo. Por ejemplo, un hombre español, de 35 años con secundaria, tendrá una probabilidad de 0,1392 de no tener trabajo si está casado frente a un 0,2147 si está soltero. Una mujer de 40 años, española y con estudios de secundaria, tendrá 9,38 puntos porcentuales más de probabilidad de no tener trabajo si es soltera frente a si está casada. En el mismo caso, pero siendo la mujer extranjera sin estudios o con estudios primarios, la probabilidad de no tener empleo será de 0,3889 en caso de estar casada frente a un 0,5182 siendo soltera, pero que se reduce hasta 0,1600 si tiene estudios superiores.

En relación a este último dato se plantea otro caso interesante, que surge de comparar a los extranjeros con diferente nivel de estudios. Para un hombre soltero, extranjero y de 30 años, la probabilidad de no tener empleo alcanza el 0,4534 si no tiene estudios o tiene estudios primarios frente a un 0,1281 si tiene estudios superiores, una cifra por debajo de la tasa de paro de la muestra (0,13) y muy parecida a individuos de las mismas características, pero de nacionalidad española, que tendrían un 0,0930.

Los estudios superiores son sin duda la característica más determinante para conseguir un empleo. Incluso si comparamos un individuo con algunas características que afectan negativamente a la probabilidad de conseguir un empleo, la probabilidad de tener trabajo es mayor si este presenta estudios superiores en varios casos. Por ejemplo, comparando dos mujeres de 40 años, una soltera y de nacionalidad extranjera (dos variables que influyen negativamente en la probabilidad de trabajar) pero con estudios superiores, con otra casada y de nacionalidad española (variables que afectan positivamente) con estudios de secundaria; observamos que la probabilidad de la primera de no trabajar es del 0,1600, frente al 0,1897 de la segunda.

Además, entre dos individuos españoles de 25 años y solteros, la probabilidad de tener un empleo asciende casi 12 puntos porcentuales en mujeres y más de 8 en hombres por el hecho de realizar estudios después de bachillerato. Este dato es muy interesante ya que representa a un gran porcentaje de jóvenes

españoles que comienzan a buscar empleo de manera activa tras la finalización de su vida académica.

5. CURVAS DE RESPUESTA

A partir de la fórmula del modelo Logit, se han realizado curvas de respuesta para estudiar la evolución en la probabilidad de encontrarse en situación de los individuos en función de su edad y distinguiendo entre cada categoría de las variables.

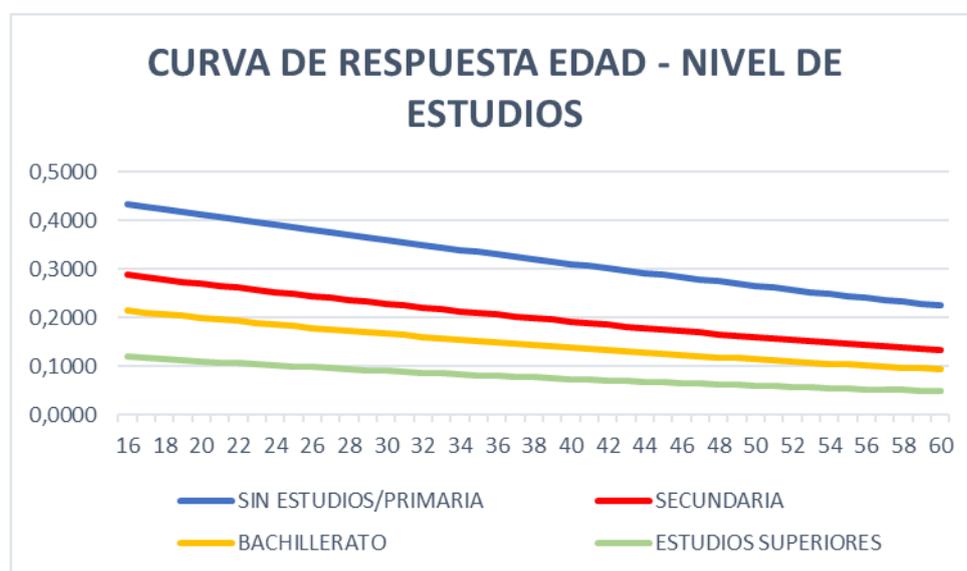


Gráfico 8 Curva de respuesta Edad - Nivel de estudios. Fuente: elaboración propia a partir de los datos del INE

En cuanto al nivel de estudios, podemos observar que la brecha entre la probabilidad de encontrarse en situación de desempleo para cada nivel de estudios se reduce, bajando especialmente esta probabilidad en individuos sin estudios o con estudios primarios, pero manteniéndose aun muy lejana del resto de individuos.

Cabe destacar que para cualquier edad los individuos con estudios superiores tienen menos de un 0,13 de probabilidad de encontrarse en desempleo.

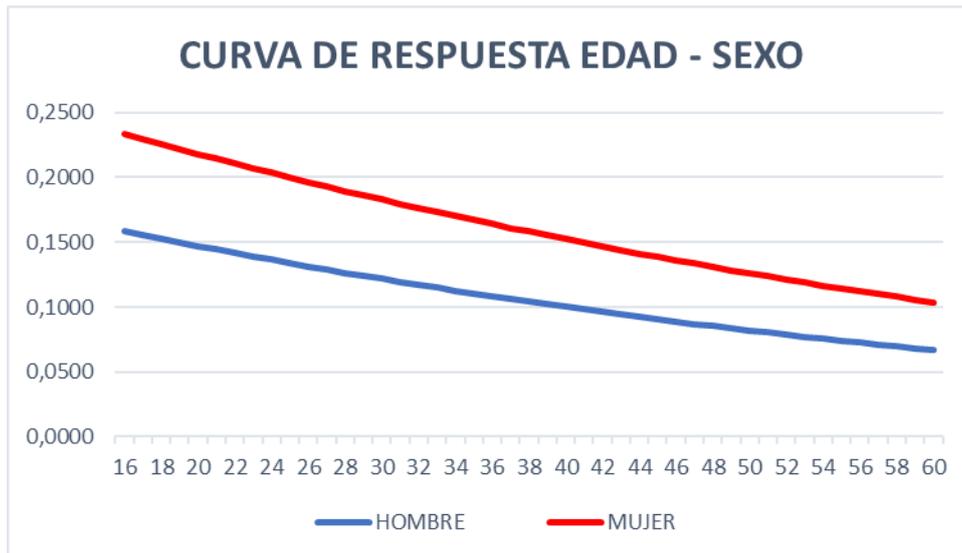


Gráfico 9 Curva de respuesta Edad - Sexo. Fuente: elaboración propia a partir de los datos del INE

La brecha entre hombres y mujeres también se reduce en individuos de mayor edad, reduciéndose de más de 7 puntos porcentuales a los 16 años a menos de 4 puntos porcentuales en individuos de 60 años.

También es interesante ver que no es hasta los 50 años que las mujeres tienen una probabilidad de estar desempleadas por debajo de la media de la muestra (13%), mientras que los hombres se encuentran por debajo de este umbral a los 28 años.

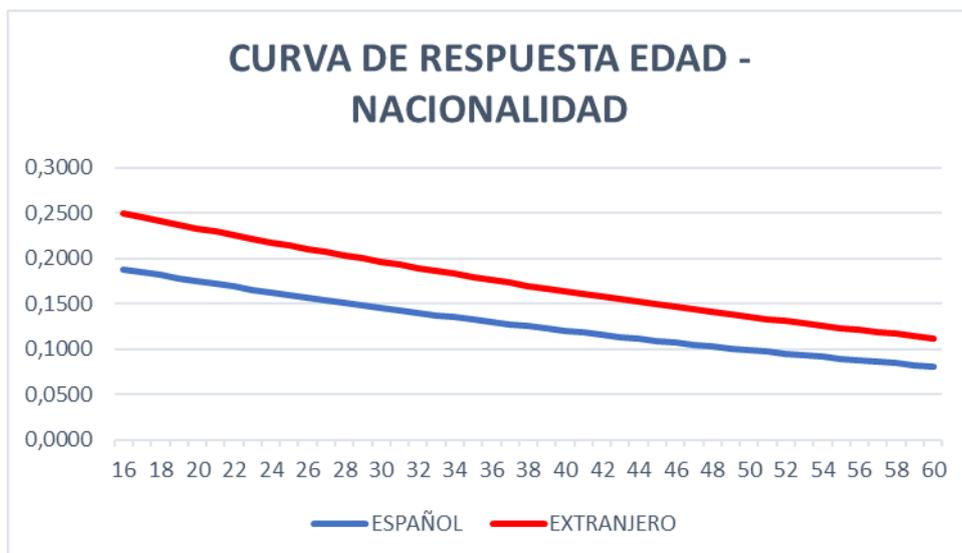


Gráfico 10 Curva de respuesta Edad - Nacionalidad. Fuente: elaboración propia a partir de los datos del INE

Para el caso de la nacionalidad encontramos resultados similares a las anteriores curvas de respuesta, reduciéndose la brecha en la probabilidad de desempleo a medida que aumentamos la edad de los individuos, pero siendo siempre inferior la probabilidad de no estar ocupado en personas de nacionalidad española. Sin embargo, en este caso ambas curvas son bastante similares en su pendiente, por lo que la diferencia entre la probabilidad de desempleo en españoles y extranjeros es bastante parecida para cualquier edad.

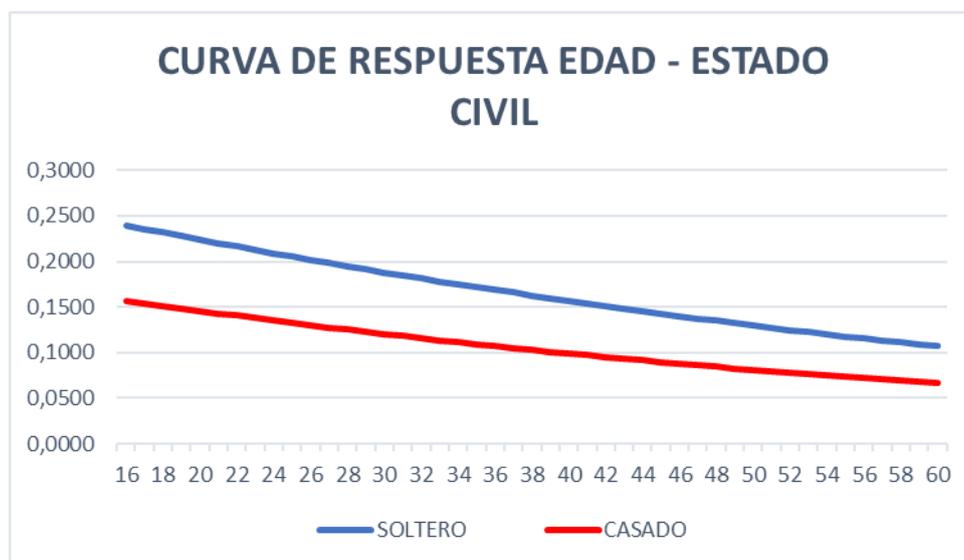


Gráfico 11 Curva de respuesta Edad - Estado civil. Fuente: elaboración propia a partir de los datos del INE

Por último, pero con conclusiones bastante similares, podemos observar que la brecha en la probabilidad de desempleo entre personas solteras y casadas también se reduce cuando aumenta la edad de los individuos, pero siendo los casados siempre los que tienen mayor probabilidad de estar ocupados.

6. CONCLUSIONES

Una vez realizados los análisis descriptivos, el análisis Logit y la estimación de escenarios; podemos extraer varias conclusiones acerca de los diferentes factores personales y su influencia en la probabilidad de encontrarse en situación de desempleo en España.

En primer lugar, hay que dejar claro que el modelo Logit refleja que todas las variables utilizadas en el estudio son significativas, por lo que todas serán útiles y representativas para concluir en la probabilidad del desempleo en individuos de perfiles muy concretos. En relación con esto, como conclusión especialmente importante hay que decir que las variables que hemos utilizado si que influyen y de hecho lo hacen en gran medida en la probabilidad de encontrarse en situación de desempleo, llegando a existir diferencias de hasta 50 puntos porcentuales en la probabilidad de no tener trabajo entre dos individuos de diferentes perfiles.

Por medio de los análisis descriptivos, hemos extraído una serie de conclusiones acerca de qué grupos de individuos sufren mayor porcentaje de desempleo dentro de cada variable. Para el caso concreto de la edad, hemos observado que el mayor porcentaje de parados se encuentra entre los jóvenes menores de 25 años, con un porcentaje considerablemente superior al resto, además, hay un descenso progresivo en el porcentaje de parados a medida que observamos segmentos de mayor edad.

Por otro lado, existe mayor porcentaje de paro entre el género femenino (en los estudios descriptivos encontramos un 11,3% en hombres frente al 14,9% en mujeres), siendo especialmente reseñable que, cuando utilizamos distinción por sexo en el estudio de cualquier variable, las mujeres siguen siendo las que mayor porcentaje de paro representan en todas las categorías. De esta manera, las mujeres españolas y extranjeras presentan más porcentaje de desempleo que hombres españoles y extranjeros respectivamente; cosa que también sucede para los estudios, dónde los hombres representan mejores tasas de ocupación para cada nivel educativo con respecto a las mujeres. Sin embargo, el único caso en el que el sexo no parece tener influencia entre personas solteras, que presentan un porcentaje de desempleo muy similar en ambos sexos.

Con relación a esta última variable, hemos observado que el estado civil repercute claramente en la tasa de paro, con un porcentaje de desempleo mayor entre los solteros (9,4% en casados y 17,2% en solteros), independientemente del sexo, ya el porcentaje de desempleo en casados es inferior al 13% tanto en hombres como en mujeres, mientras que, en solteros, el porcentaje de parados supera el 17% en ambos sexos.

En cuanto a la nacionalidad, son los extranjeros los que presentan un peor porcentaje de ocupación, sin embargo, el número total de encuestados corresponde a menos del 10% de la muestra, por lo que el desempleo entre personas de nacionalidad española sigue siendo muy similar al total de la muestra.

Con respecto al nivel de estudios, era de esperar que los niveles más altos de educación correspondiesen a las menores tasas de paro, pero cabe destacar la importancia tan significativa que tienen los estudios superiores, cuyos individuos pertenecientes a este grupo presentan un 92,3% de ocupación, siendo el único grupo por encima del porcentaje global de la muestra.

Los resultados comentados de los análisis descriptivos han sido respaldados posteriormente por el análisis Logit, que cuantificado la influencia real de las categorías de cada variable en la probabilidad de encontrarse en situación de desempleo.

En el análisis Logit, hemos observado de nuevo la gran importancia de los estudios superiores, una variable que tiene un Odd Ratio de 0,177, por lo que poseer este tipo de estudios reduce el cociente entre la probabilidad de estar desempleado y no estarlo un 82,3% con respecto a los individuos sin estudios o con estudios primarios. Además, en los diferentes escenarios que hemos planteado, los diferentes perfiles de individuos que hemos simulado con esta categoría presentan tasas de paro considerablemente inferiores a otros de mismas características, pero menos formados educativamente.

De la misma manera, podemos decir que los extranjeros, especialmente los solteros, tienen una probabilidad muy alta de no tener empleo, con excepción de aquellos que tienen educación superior, cuya probabilidad de ocupación es realmente buena y muy similar al de individuos españoles.

Cabe destacar que hemos respaldado también los análisis descriptivos del resto de variables (edad, sexo y estado civil), cuyos resultados tras la estimación del modelo Logit y planteamiento de los escenarios ha reafirmado que, manteniendo el resto de variables idénticas, los hombres tienen una menor probabilidad de desempleo (con un Odd Ratio de 0,619) así como las personas casadas (con un Odd Ratio de 1,690); y los jóvenes son los que mayor probabilidad tienen de no

tener empleo, especialmente los españoles, siendo la diferencia entre jóvenes e individuos de otros grupos de edad mucho más amplia que en los extranjeros, cuya probabilidad de desempleo se corresponde más bien al estado civil y nivel de estudios.

En el método de los escenarios, hemos obtenido la menor probabilidad de encontrarse en situación de desempleo posible, que es de 0,0304, ya que se ha planteado un individuo de un perfil en el que todas las variables afectan positivamente a tener un empleo (hombre español de 60 años, casado y con estudios superiores).

Por otro lado, el individuo con mayor probabilidad de encontrarse en situación de desempleo es una mujer extranjera de 40 años, soltera y con estudios primarios (o sin estudios), con un 0,5182 de probabilidad de no tener trabajo, un dato que aun podría ser inferior si la mujer tuviera 16 años (el mínimo de la muestra).

Por último, hemos contrastado por medio de las curvas de respuesta que la edad de los individuos reduce la brecha entre las distintas categorías de cada variable, existiendo menor diferencia en la probabilidad de encontrarse en situación de desempleo para todos los individuos a medida que aumenta la edad.

7. BIBLIOGRAFÍA

«El paro de la eurozona continúa en mínimos históricos con un 6,5%. España continúa a la cabeza de la lista, con una tasa del 12,4%». Cinco Días. 9 de enero de 2023. Consultado en mayo de 2023.

https://cincodias.elpais.com/cincodias/2023/01/09/economia/1673259412_047689.html

Sanchez-Silvia, Carmen. “Historia de un paro que no cesa” – El País. 20 de noviembre de 2015. Consultado en junio de 2023.

https://elpais.com/economia/2015/10/29/actualidad/1446133696_808163.html

Centro de Investigaciones sociológicas. «Barómetro del CIS marzo 2022». Consultado en junio de 2023.

https://www.cis.es/cis/export/sites/default/-Archivos/Marginales/3340_3359/3355/es3355mar.html

Cuadrado Roura, Juan Ramón. “La terciarización de la economía española: algunos rasgos estilizados del período 1975-2014.” 30 de junio de 2016. Consultado en mayo de 2023.

<https://revistasice.com/index.php/ICE/article/view/1852>

Eurostat: “Total unemployment rate”. Consultado en junio de 2023.

<https://ec.europa.eu/eurostat/web/products-datasets/-/tps00203>

Goldberger, A.S. (2001): Introducción a la Econometría, Ariel, Barcelona.

INE - Instituto Nacional de Estadística: “Encuesta de Población Activa T1 2022”. Consultado en mayo de 2023.

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

Organización para la Cooperación y el Desarrollo Económicos: “Desempleo en España”. Consultado en mayo de 2023.

<https://www.oecd.org/espanol/estadisticas/tasa-desempleo.htm>

Prieto Alaiz, María Mercedes (2021) Apuntes de clase de la asignatura de Análisis Multivariante de Datos Económicos del grado de Marketing e Investigación de Mercados. Universidad de Valladolid.

Vega Crespo, Josefa (2020) Apuntes de clase de la asignatura de Estructura Económica Sectorial del grado de Marketing e Investigación de Mercados. Universidad de Valladolid.

ANEXO

Se adjuntan a continuación las tablas originales completas obtenidas en SPSS para el análisis descriptivo.

Tabla cruzada EDAD2*AOI1

			AOI1		Total
			Ocupado	Desocupado	
EDAD2	16 a 24 años	Recuento	2519	1144	3663
		% dentro de EDAD2	68,8%	31,2%	100,0%
	25 a 44 años	Recuento	20566	3106	23672
		% dentro de EDAD2	86,9%	13,1%	100,0%
	45 a 59 años	Recuento	23977	2950	26927
		% dentro de EDAD2	89,0%	11,0%	100,0%
	60 o más años	Recuento	5440	668	6108
		% dentro de EDAD2	89,1%	10,9%	100,0%
Total		Recuento	52502	7868	60370
		% dentro de EDAD2	87,0%	13,0%	100,0%

Tabla cruzada SEXO1*AOI1

			AOI1		Total
			Desocupado	Ocupado	
SEXO1	Varón	Recuento	3523	27714	31237
		% dentro de SEXO1	11,3%	88,7%	100,0%
	Mujer	Recuento	4345	24788	29133
		% dentro de SEXO1	14,9%	85,1%	100,0%
Total		Recuento	7868	52502	60370
		% dentro de SEXO1	13,0%	87,0%	100,0%

Tabla cruzada NFORMA1*AOI1

			AOI1		Total
			Desocupado	Ocupado	
NFORMA1	Sin estudios	Recuento	187	424	611
		% dentro de NFORMA1	30,6%	69,4%	100,0%
	Educación primaria	Recuento	605	1546	2151
		% dentro de NFORMA1	28,1%	71,9%	100,0%
	Educación secundaria	Recuento	3014	14039	17053
		% dentro de NFORMA1	17,7%	82,3%	100,0%
	Bachillerato o equivalente	Recuento	2059	12356	14415
		% dentro de NFORMA1	14,3%	85,7%	100,0%
	Estudios superiores	Recuento	2003	24137	26140
		% dentro de NFORMA1	7,7%	92,3%	100,0%
Total	Recuento		7868	52502	60370
	% dentro de NFORMA1		13,0%	87,0%	100,0%

Tabla cruzada NFORMA1*AOI1*SEXO1

			AOI1		Total	
			Desocupado	Ocupado		
Varón	NFORMA1	Sin estudios	Recuento	102	283	385
			% dentro de NFORMA1	26,5%	73,5%	100,0%
		Educación primaria	Recuento	330	986	1316
			% dentro de NFORMA1	25,1%	74,9%	100,0%
		Educación secundaria	Recuento	1455	8775	10230
			% dentro de NFORMA1	14,2%	85,8%	100,0%
		Bachillerato o equivalente	Recuento	899	6631	7530
			% dentro de NFORMA1	11,9%	88,1%	100,0%
		Estudios superiores	Recuento	737	11039	11776
			% dentro de NFORMA1	6,3%	93,7%	100,0%
Total	Recuento		3523	27714	31237	
	% dentro de NFORMA1		11,3%	88,7%	100,0%	
Mujer	NFORMA1	Sin estudios	Recuento	85	141	226
			% dentro de NFORMA1	37,6%	62,4%	100,0%
		Educación primaria	Recuento	275	560	835
			% dentro de NFORMA1	32,9%	67,1%	100,0%
		Educación secundaria	Recuento	1559	5264	6823
			% dentro de NFORMA1	22,8%	77,2%	100,0%
		Bachillerato o equivalente	Recuento	1160	5725	6885
			% dentro de NFORMA1	16,8%	83,2%	100,0%
		Estudios superiores	Recuento	1266	13098	14364
			% dentro de NFORMA1	8,8%	91,2%	100,0%
Total	Recuento		4345	24788	29133	
	% dentro de NFORMA1		14,9%	85,1%	100,0%	

Tabla cruzada NAC2*AOI1

			AOI1		Total
			Desocupado	Ocupado	
NAC2	Extranjero	Recuento	982	3470	4452
		% dentro de NAC2	22,1%	77,9%	100,0%
	Español	Recuento	6886	49032	55918
		% dentro de NAC2	12,3%	87,7%	100,0%
Total	Recuento		7868	52502	60370
	% dentro de NAC2		13,0%	87,0%	100,0%

Tabla cruzada NAC2*AOI1*SEXO1

				AOI1		Total	
				Desocupado	Ocupado		
SEXO1	Varón	NAC2	Extranjero	Recuento	426	1849	2275
				% dentro de NAC2	18,7%	81,3%	100,0%
		Español	Recuento	3097	25865	28962	
			% dentro de NAC2	10,7%	89,3%	100,0%	
		Total		Recuento	3523	27714	31237
	% dentro de NAC2		11,3%	88,7%	100,0%		
	Mujer	NAC2	Extranjero	Recuento	556	1621	2177
				% dentro de NAC2	25,5%	74,5%	100,0%
		Español	Recuento	3789	23167	26956	
			% dentro de NAC2	14,1%	85,9%	100,0%	
Total		Recuento	4345	24788	29133		
% dentro de NAC2		14,9%	85,1%	100,0%			

Tabla cruzada ECIV2*AOI1

			AOI1		Total
			Desocupado	Ocupado	
ECIV2	Casado	Recuento	3044	29305	32349
		% dentro de ECIV2	9,4%	90,6%	100,0%
	Soltero	Recuento	4824	23197	28021
		% dentro de ECIV2	17,2%	82,8%	100,0%
Total	Recuento		7868	52502	60370
	% dentro de ECIV2		13,0%	87,0%	100,0%

Tabla cruzada ECIV2*AOI1*SEXO1

SEXO1				AOI1		Total
				Desocupado	Ocupado	
Varón	ECIV2	Casado	Recuento	1111	15925	17036
			% dentro de ECIV2	6,5%	93,5%	100,0%
	Soltero	Recuento	2412	11789	14201	
		% dentro de ECIV2	17,0%	83,0%	100,0%	
	Total	Recuento	3523	27714	31237	
		% dentro de ECIV2	11,3%	88,7%	100,0%	
Mujer	ECIV2	Casado	Recuento	1933	13380	15313
			% dentro de ECIV2	12,6%	87,4%	100,0%
	Soltero	Recuento	2412	11408	13820	
		% dentro de ECIV2	17,5%	82,5%	100,0%	
	Total	Recuento	4345	24788	29133	
		% dentro de ECIV2	14,9%	85,1%	100,0%	