



Universidad de Valladolid

Facultad de Ciencias
Económicas y Empresariales

Trabajo de Fin de Grado

Grado en Marketing e Investigación de
Mercados

ANALISIS DE LOS DETERMINANTES
DEL DESEMPLEO EN ESPAÑA: TERCER
TRIMESTRE DE 2019

Presentado por:

Rocío Peidro Morillo

Valladolid, 21 de Julio de 2020

RESUMEN

El desempleo, entendido como la situación en la que un individuo que se encuentra en condiciones de trabajar no tiene empleo o lo ha perdido, es una de las situaciones que más preocupa a nivel general, puesto que un país con un alto nivel de desempleo da a entender que su gobierno ejerce una mala gestión. En los últimos años, la tasa de paro ha disminuido a nivel mundial tras recuperarse de la crisis de la burbuja inmobiliaria. Partiendo de los resultados obtenidos en la Encuesta de Población Activa del 3^{er} trimestre de 2019, se ha realizado este trabajo basado en un análisis econométrico aplicando un modelo de regresión logística binaria, que permite conocer el efecto de una serie de variables que presenta el individuo que hace que aumenten o disminuyan las probabilidades de encontrarse desempleado. Los resultados se han obtenido a través del programa informático Eviews, dando lugar a que las variables seleccionadas (edad, sexo, nacionalidad, estado civil y nivel de estudios) presentan una fuerte influencia sobre el empleo en España.

Palabras clave: desempleo, modelo Logit, encuesta de población activa.

Clasificación JEL: C25, C51, E24

ABSTRACT

Unemployment as the situation where a person is in a condition to work and doesn't work or he lost it. It's one of the most worried situation at the World, because a country with a high unemployment rate means that the government makes a bad gestation. In recent years, the unemployment rate has declined globally after recovering from the estate bubble. As the results obtained in the Labor Force Survey of the 3rd quarter of 2019, this project was done based on an econometric analysis applying a binary logistic regression model, it allows to know the effect of amount of variables who a person has and makes the probability of being unemployed increase or not. The final results were obtained by Eviews and showing that the selected variables (age, sex, nationality, marital status and education level), have a strong influence in Spain about the employment data.

Key words: Unemployment, logistic regression model, Labor Force Survey.

Índice

1. INTRODUCCIÓN.....	5
2. EL DESEMPLEO Y SU EVOLUCIÓN.....	6
2.1. Desde 1950 hasta 1976	8
2.2. Desde 1976 hasta 1985	9
2.3. Desde 1986 hasta 1996	9
2.4. Desde 1997 hasta 2007	9
2.5. Desde 2008 hasta 3er trimestre de 2019	9
2.6. 3er trimestre de 2019.....	11
3. METODOLOGÍA.....	12
3.1. Regresión de Probabilidad Lineal.....	12
3.2. Modelos no lineales.....	13
3.2.1. <i>Modelo Logit</i>	14
3.2.2. <i>Modelo Probit</i>	17
4. BASE DE DATOS.....	18
5. VARIABLES UTILIZADAS	19
5.1. Situación Laboral	19
5.2. Edad.....	19
5.3. Sexo.....	20
5.4. Nacionalidad.....	20
5.5. Estado Civil.....	20
5.6. Nivel de estudios.....	21
6. RESULTADOS	21
6.1. Estimación y contrastes.....	22
6.2. Bondad del ajuste	24
6.3. El efecto marginal de las variables explicativas.....	25
6.4. Escenarios	28
6.5. Curvas de respuesta	32
7. CONCLUSIONES	35
8. BIBLIOGRAFÍA.....	38
ANEXO I. ESCENARIOS.....	40

ÍNDICE DE TABLAS

Tabla 2.1. Porcentaje de desempleo por sexo en 2019 en España.....	11
Tabla 6.1. Estimación del modelo Logit.....	15
Tabla 6.2. Porcentaje de predicciones correctas.....	18
Tabla 6.3. Efectos marginales de las variables cualitativas del modelo.....	21
Tabla 6.4. Escenarios de probabilidad.....	31

ÍNDICE DE GRÁFICOS

Gráfico 2.1. Evolución del trabajo en España.....	7
Gráfico 2.2. Encuesta de Población Activa.....	10
Gráfica 6.1 Curva de respuesta de la edad en función del sexo.....	34
Gráfica 6.2 Curva de respuesta de la edad en función del estado civil.....	34
Gráfica 6.3 Curva de respuesta de la edad en función de la nacionalidad.....	35
Gráfica 6.4 Curva de respuesta de la edad en función del nivel de estudios...	35

1. INTRODUCCIÓN

Durante años, el desempleo ha sido una de las grandes preocupaciones de la población de todo el mundo. Gracias a los datos que recogen distintas organizaciones sociales y económicas como la Seguridad Social o el Instituto Nacional de Estadística, tanto los Gobiernos como la gente de a pie puede acceder a la información sobre la situación en la que se encuentra su país en este tema. Con ello, se puede conocer la evolución que se ha ido produciendo en España desde que se tienen datos.

A principios del año 2020, se hizo pública la encuesta del CIS (Centro de Investigaciones Sociológicas) realizada en Diciembre de 2019 donde se exponían las principales preocupaciones de la población española, siendo el paro la principal de las preocupaciones, con un apoyo del 57,4% de los encuestados. Este porcentaje ha disminuido en relación a la anterior encuesta que se realizó en noviembre del mismo año, donde el 60,3% de la población apoyaba el paro como la mayor de sus preocupaciones, pudiendo deberse a dos opciones: la primera es la posibilidad de haber encontrado empleo o haberse jubilado, y la segunda es que otras situaciones externas han adquirido una mayor relevancia (situación política, enfermedades, terrorismo...)

En España la tasa de desempleo actual es del 14,5%, aunque es un valor muy alto, ha disminuido desde la última crisis económica que comenzó en 2008. El gobierno tiene que seguir luchando para fomentar el empleo y proteger a los trabajadores del despido.

Pero, ¿cuáles son las variables que producen que un individuo tenga más probabilidades de encontrarse desempleado? Este trabajo tiene como objetivo contestar a esa pregunta utilizando determinadas variables como la edad, el sexo y el nivel de estudios entre otras. Por tanto, con las técnicas que se van a utilizar a lo largo del estudio, se determinará la probabilidad estimada que tiene un individuo, con una serie de características determinadas, de encontrarse en el grupo de desempleados.

La estimación planteada, teniendo en cuenta las características del desempleo, se realizará a través de un modelo de elección discreta, más concretamente, un modelo Logit. Para una mejor comprensión de este modelo, en el tercer

apartado del trabajo se encuentran todas las características principales, y se utilizarán los datos proporcionados por la Encuesta de Población Activa (EPA) del 3^{er} trimestre de 2019, que serán analizados con el programa estadístico Eviews.

El trabajo está estructurado en seis secciones. En la primera se analiza el desempleo para poder conocer este concepto y su evolución desde que se obtienen datos. A continuación, se analiza el modelo Logit de forma general para entender la técnica utilizada con los datos y, por tanto también es necesario conocer la base de datos con los que se va a trabajar. Posteriormente se explica de forma detallada la construcción e interpretación del modelo y los resultados obtenidos. Para finalizar, se extraen las conclusiones de los diferentes análisis realizados con el fin de contestar a la pregunta que define el objetivo de este trabajo.

2. EL DESEMPLEO Y SU EVOLUCIÓN

Primero, debemos saber a qué nos referimos con desempleo, según el INE son las personas de 16 años o más que durante la semana de referencia han estado sin trabajo, disponibles para trabajar y buscando activamente empleo. Se considera que una persona busca empleo de forma activa si:

- Ha estado en contacto con una oficina pública de empleo con el fin de encontrar trabajo.
- Ha estado en contacto con una oficina privada (oficina de empleo temporal, empresa especializada en contratación, etc.) con el fin de encontrar trabajo.
- Ha enviado una candidatura directamente a los empleadores.
- Ha indagado a través de relaciones personales, por mediación de sindicatos, etc.
- Se ha anunciado o ha respondido a anuncios de periódicos.
- Ha estudiado ofertas de empleo.
- Ha participado en una prueba, concurso o entrevista, en el marco de un procedimiento de contratación.
- Ha estado buscando terrenos o locales.

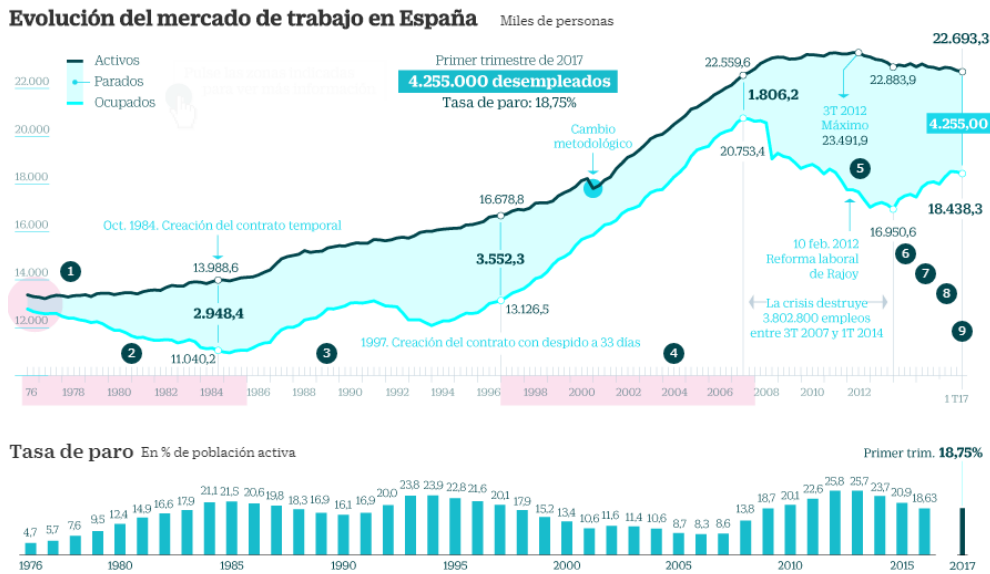
- Ha realizado gestiones para obtener permisos, licencias o recursos financieros.

También se consideran parados a aquellos individuos que han conseguido un puesto de trabajo pero aún no se han incorporado a él.

El desempleo empezó a contabilizarse en España en 1964 a través de la Encuesta de Población Activa (EPA), se trata de una investigación que se realiza de forma continuada con una periodicidad de 3 meses (trimestral) que está dirigida a las familias. Tiene como fin la recogida de datos del empleo tanto para ocupados y parados como para los inactivos.

La evolución del desempleo desde que empezó a contabilizarse hasta hoy ha pasado por muchos altibajos principalmente relacionados con el ciclo económico en el que se encontraba el país. A continuación se va a realizar un análisis del desempleo desde los años 50 hasta hoy a rasgos generales.

En primer lugar, se puede ver una evolución del mercado de trabajo en España, desde 1976 hasta el primer trimestre de 2017. En el gráfico 2.1, se puede observar la línea de población activa, la línea de población ocupada, y entre medias de ambas, el grupo de individuos que se encuentra desempleado a lo largo de toda la línea temporal. Gracias a ellas se puede determinar cuando existe un mayor o menor tasa de paro en el país.



¹ Gráfico 2.1: Evolución del trabajo en España

2.1. Desde 1950 hasta 1976

Aunque durante este periodo aún no se realizaban encuestas sobre el desempleo, si se tienen datos de como era el estado del empleo en ese momento.

El desempleo era muy bajo, pero debido a la incapacidad productiva por la que estaba pasando el país, hizo que el paro aumentase. El plan de Estabilización aprobado en 1959, supuso el cierre de empresas.

Durante este periodo, la población activa disminuyó, puesto que el número de años que se dedicaban a los estudios aumentó, y por tanto se incorporaron más tarde al mercado laboral, además no contar con la mujer trabajadora, que trabajaba en negro.

Por otro lado, la emigración del campo a la ciudad supone un descenso en el desempleo de los sectores secundarios y terciarios, y con ello aumentó el empleo femenino al incorporarse las mujeres al mercado laboral de manera oficial.

¹ Gráfico 2.1: Evolución del mercado de trabajo en España: https://cincodias.elpais.com/cincodias/2014/07/28/graficos/1406544661_293795.html

2.2. Desde 1976 hasta 1985

Durante este período se consolidó la incorporación de la mujer al mundo laboral, dando lugar a una disminución del paro femenino y aumentando el masculino. Durante este periodo el paro también se vio golpeado por la crisis del petróleo, la subida de los salarios y unos altos tipos de interés, llegando a finalizar este período con un porcentaje de población desocupada del 21,5%.

2.3. Desde 1986 hasta 1996

Hasta 1990, por fortuna, el sector de la construcción se encontraba en un buen nivel de actividad. Se crearon 1 millón de empleos, pero esta situación no perduró en el tiempo. Debido a los altos niveles de gasto público que soportaba el país, se produjo un aumento del déficit. Fue entonces cuando entre 1990 y 1994 se pierde todo el empleo que se había creado hasta el momento, llegando a un nivel de desempleo superior al de 1974.

2.4. Desde 1997 hasta 2007

Este periodo se centró sobretodo en el desempleo juvenil, situándose en un 35% (antes del comienzo de la crisis económica). En estos años, muchos jóvenes decidieron abandonar los estudios para solicitar empleos no cualificados ya que se habían creado 5,5 millones de puestos en los que no se requería ninguna formación o ésta era mínima, lo que dio lugar a que se consiguiera en la primavera del 2007 el valor más bajo en la historia de España en paro.

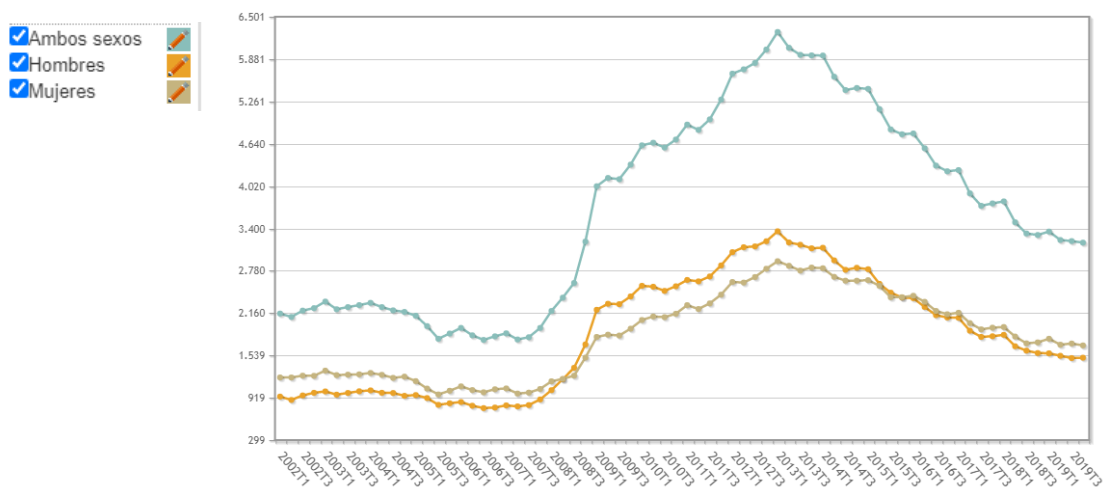
2.5. Desde 2008 hasta 3er trimestre de 2019

En el año 2008, se produjo el estallido de la burbuja inmobiliaria en España. Fue un año duro tanto económicamente como a nivel social, puesto que muchos puestos de trabajo se vieron destruidos debido a las circunstancias que el país estaba sufriendo, pero no fue hasta 2013 cuando se produjo el máximo histórico registrado, llegando a alcanzar un 27,16%. Esto vino motivado por los primeros efectos de la instauración de la Reforma Laboral en Febrero de 2012 por el Gobierno.

Durante los años 2009 hasta 2013 el porcentaje de paro no dejó de aumentar, en 2014 comenzó a registrarse un descenso en el desempleo, pero no significó que aumentase la ocupación, si no que el número de individuos de población

activa descendió. El año 2016, hasta el día de hoy, es el que presenta el mejor dato de recuperación con un descenso del paro de 305.000 personas.

En 2018, la tasa de desempleo se encontraba por debajo del 15%, algo que no ocurría desde 2008.



² Gráfico 2.2: La evolución del paro por sexos. (Fuente: INE)

En el gráfico 2.2, se puede ver cómo afecta el paro a hombres y mujeres por separado y en conjunto de entre 16 a más de 70 años durante el 2002 hasta el 2º trimestre de 2019. Como se puede apreciar, en el año 2008 (momento en el que explotó la burbuja inmobiliaria), los niveles de paro fueron en aumento de manera muy pronunciada y se mantuvo así hasta el 1º trimestre de 2013, donde el índice de paro empezó a disminuir, pero aun así los niveles seguían siendo muy altos. Las mujeres se han visto más afectadas que los hombres en cuanto al nivel de desempleo, excepto durante la recesión, que hombres y mujeres sufrieron el paro de manera muy similar. Ahora que comienza a mejorar la situación laboral del país, vuelve a hacerse notar una distancia entre los niveles de un sexo con el otro.

A lo largo de este trabajo se conocerán cuáles de las variables seleccionadas para el modelo son más influyentes en el momento en el que un individuo se encuentra en una situación laboral de desempleo.

² Gráfico 2.2 Encuesta de Población Activa. Fuente: Instituto Nacional de Estadística (INE): <https://www.ine.es/jaxiT3/Datos.htm?t=4086>

2.6. 3er trimestre de 2019

Anteriormente, se ha puesto en conocimiento la evolución del desempleo desde que se tienen datos en España. Se va a estudiar un periodo concreto, el 3^{er} trimestre de 2019, para lo cual se exponen los datos disponibles:

	TOTAL
HOMBRES	12,17%
MUJERES	15,92%
AMBOS	13,92%

³ Tabla 2.1: Porcentaje de desempleo por sexo 3º trimestre de 2019 en España

Con esta tabla, se puede apreciar la diferencia que existe entre el paro de los hombres y el de las mujeres, dando a entender que al ser hombre tienes menos probabilidades de encontrarte en paro que siendo mujer. Actualmente el nivel de paro se encuentra en casi un 14%, de lo que se deduce que ha descendido con respecto al trimestre anterior, que se encontraba en un 14,02%.

En los últimos años, en España el empleo ha sido un tema muy importante, puesto que durante la crisis económica que sufrió el país, muchos individuos se vieron sin trabajo y subsistiendo a través de diferentes ayudas, bien por parte del Estado (subsidio de desempleo), ayuda de sus familiares y/o amigos. Por lo que se puede determinar que el desempleo está muy ligado al ciclo económico que atraviese el país, de forma que cuando este se encuentra en recesión, los niveles de empleo caen y al contrario cuando se encuentra en auge.

³ Tabla 2.1. Porcentaje de paro por sexo a nivel nacional, (INE): <https://www.ine.es/jaxiT3/Datos.htm?t=4086>

3. METODOLOGÍA

El objetivo de este trabajo es estudiar los determinantes del desempleo en España durante el 3^{er} trimestre de 2019, concretamente el grado de influencia que tienen las diferentes características seleccionadas sobre la probabilidad de que un individuo se encuentre o no desempleado. Para comenzar, se va a realizar un análisis de los distintos tipos de análisis multivariantes que se utilizan teniendo una variable endógena binaria, entre ellos se encontrará el análisis Logit, que es el que se va a utilizar.

La variable independiente es la que proporciona la información necesaria para determinar el uso de un tipo de análisis u otro, esta variable puede presentarse de manera cuantitativa o cualitativa. Las primeras son aquellas en las que la variable puede tomar cualquier valor sin ninguna restricción, mientras que las segundas es el investigador quien les asigna un valor. Puesto que son estas últimas las se van a estudiar, a continuación se analizarán los dos posibles casos: lineal (Regresión de probabilidad lineal), y no lineal (análisis Logit y análisis Probit).

Los factores cualitativos principalmente aportan información binaria (sexo, situación laboral, etc.), para ello se trabaja con variables binarias o dummy. Estas variables solo admiten dos valores, el cero y el uno, al definir una variable se debe decidir que característica se le asigna a cada uno, por ejemplo 1 = mujeres, 0 = hombres. Asignar un valor u otro a una determinada característica no es determinante en el resultado, simplemente se escoge para tener una mejor comprensión del modelo y su interpretación es más sencilla.

3.1. Regresión de Probabilidad Lineal

Los modelos de regresión de probabilidad lineal (MRPL) son aquellos donde la variable endógena es una variable binaria, toma el valor 1 cuando ocurre el objeto de estudio y 0 cuando no, y se encuentra en función de las X (variables explicativas). El modelo se presentaría:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \varepsilon_i$$

Se debe tener en cuenta que $E(Y|X) = X_i'\beta$, suponiendo que $E(\varepsilon_i) = 0$. Sabiendo que (P) es la probabilidad de que $Y = 1$, y (1-P) la probabilidad de que ocurra lo contrario, $Y = 0$, la función sería:

$$f(y_i) = P_i^y (1 - P_i)^{1-y} \text{ para } Y=1,0 \rightarrow E(Y_i) = 0(1-P_i) + 1(P_i) = P_i$$

La clave del análisis se encuentra siempre que $E(Y_i) = E(Y|X)$, es decir, que la esperanza condicionada de Y_i , se puede interpretar como la probabilidad condicionada de Y_i : $E(Y|X) = P_i$.

Este modelo se estima utilizando mínimos cuadrados ordinarios (MCO), en este caso se interpreta como las variaciones en la probabilidad de éxito dado un aumento en una unidad una de las variables explicativas.

Los MRPL presentan tres desventajas fundamentales:

- Las probabilidades ajustadas pueden ser inferiores a 0 o superiores a 1.
- Las perturbaciones no siguen una distribución normal.
- Presenta heterocedasticidad.

3.2. Modelos no lineales

El análisis Logit y Probit, son modelos no lineales y ambos presentan muchas similitudes y pocas diferencias, por ello, primero se analizará el análisis Logit con todos los detalles, puesto que es el modelo de estudio de este trabajo, y posteriormente se analizará el Probit, donde se mostrará únicamente las diferencias que presenta frente al modelo Logit.

En primer lugar, la ventaja de utilizar un modelo no lineal es que las probabilidades ajustadas se encuentran entre cero y uno, mientras que la desventaja que presentan estos modelos es que su interpretación es difícil.

3.2.1. Modelo Logit

Un modelo logit es un modelo de elección discreta binaria que permite explicar una situación en términos de probabilidad. El modelo lineal presenta limitaciones, por ello, para poder eliminarlas, se consideran una clase de modelos de respuesta binaria, explicada de la siguiente manera:

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

Dónde F solo admite valores entre el cero y el uno. $0 \leq F \leq 1$, para todos los número reales de X. Con esto se puede asegurar que las probabilidades de respuesta se encuentren entre el intervalo 0,1.

En el modelo Logit, F es la función de distribución acumulada para una logística: $F(x) = \frac{e^x}{1+e^x}$, que está entre 0 y 1 para todos los números reales y es creciente. Esta es la función de distribución acumulada para una variable aleatoria logística estándar.

En la mayoría de las aplicaciones de los modelos de respuesta binaria, el objetivo principal es explicar los efectos de los regresores (X_{ki}) sobre la probabilidad de respuesta, $P(Y=1|X)$.

Para calcular el efecto marginal de las variables continuas, es necesario calcular la derivada parcial de X_j :

$$\frac{\partial p_i}{\partial X_{ij}} = f(X_i' \beta) \beta_j$$

$f(X_i' \beta) \beta_j \rightarrow$ Función de densidad de probabilidad

La función de distribución es creciente, y por tanto $f(x) > 0$ para todas las X. Por tanto, el efecto marginal de X_j sobre la probabilidad depende de X a través de la cantidad positiva $f(X_i' \beta) \beta_j$, lo que significa que el efecto marginal siempre tiene el mismo signo de β . Esto quiere decir por ejemplo, que si el coeficiente de la variable edad resulta positivo, a mayor edad, mayor sería la probabilidad de estar en paro. Por tanto, β_j determina el sentido (la magnitud es determinada por las observaciones).

Para estimar los modelos lineales, se pueden usar dos métodos: mínimos cuadrados ordinarios (MCO) o ponderados (MCP). Para los modelos no lineales, se podría utilizar una versión no lineal de los mínimos cuadrados (ordinarios o ponderados), pero es preferible utilizar un tercer método más acertado denominado estimador por máxima verosimilitud (EMV). El EMV escogerá como valor estimado, aquel que tenga mayor probabilidad de ocurrir según los datos observados. Está basado en la distribución de Y dada X.

La EMV sigue la siguiente distribución:

$$Y_i = \begin{cases} 1 \rightarrow P_i = P(Y_i=1) = F(X_i' \beta) & \text{Probabilidad de que ocurra} \\ 0 \rightarrow 1 - P_i = P(Y_i=0) = 1 - F(X_i' \beta) & \text{Probabilidad de que no ocurra} \end{cases}$$

La función de densidad de Y_i dada X_i , se formula:

$$f(Y_i) = [F(X_i' \beta)]^Y [1 - F(X_i' \beta)]^{1-Y}; Y = 0 \text{ o } Y = 1$$

Dónde se puede ver con facilidad que cuando $Y = 1$, se obtiene $F(X_i' \beta)$, y cuando $Y = 0$, se obtiene $1 - F(X_i' \beta)$.

La función de verosimilitud es:

$$L(\beta) = \prod_{i=1}^n f(Y_i) = \prod_{i=1}^n [F(X_i' \beta)]^Y [1 - F(X_i' \beta)]^{1-Y}$$

La función de log-verosimilitud para las observaciones i es una función de los parámetros y los datos (X_i, Y_i) , y se obtiene al aplicar logaritmos a la ecuación anterior, y por tanto, la ecuación sería:

$$\ln(L(\beta)) = \sum_{i=1}^n [Y_i \ln(F(X_i' \beta)) + (1 - Y_i) \ln(1 - F(X_i' \beta))]$$

La EMV de β denotada como $\hat{\beta}_{MV}$ maximiza esta log-verosimilitud.

Los coeficientes proporcionan los signos de los efectos marginales de cada regresor sobre la probabilidad de respuesta, y la significación estadística de X está determinada por si se puede rechazar la hipótesis nula ($H_0: \beta = 0$) a un

nivel de significación pequeño, en este estudio se va a trabajar con un nivel de significación del 5%.

En estos modelos no lineales, se puede conocer la bondad del ajuste de dos maneras diferentes, una a través del pseudo R-cuadrado de McFadden (1974) y la otra a través del porcentaje correctamente predicho. Aunque esta medida es útil, también puede ser confusa. En algunas ocasiones ocurre que el modelo predice muy bien una situación, pero muy mal la contraria, es por eso que lo mejor es conocer los tres porcentajes de acierto, cuando $Y = 0$, $Y = 1$ y el porcentaje global, y con ello se obtendrá la capacidad predictiva del modelo. Es el análisis que se va a utilizar posteriormente.

		Estimado		Porcentaje correcto
		$\hat{p}_i \geq 0.5$	$\hat{p}_i < 0.5$	
Observado	$Y_i = 1$	Acierto	Error	%
	$Y_i = 0$	Error	Acierto	%
	Global			%

El pseudo R^2 de McFadden (1974), es una medida no se utilizará en este trabajo, pero sí que se va a comentar para conocerlo. Se utiliza debido a que la R^2 común no es compatible con los modelos de elección discreta. Sus valores oscilan entre 0 y 1.

Por el modelo de pseudo R^2 de McFadden, se puede conocer la relación que existe entre el logaritmo de la función de verosimilitud del modelo estimado, con la totalidad de las variables explicativas (L_{sr}), y el modelo restringido, en el que los valores de los parámetros son cero, excepto el de la constante (L_{cr}).

$$Pseudo R^2 = 1 - \frac{\ln(L_{sr})}{\ln(L_{cr})}$$

En un sentido estricto del significado de bondad del ajuste, no es una medida, debido a que este tipo de modelos no puede explicar la variación de los valores de Y respecto a su media, ya que sus variables no tienen sentido, son categorías o etiquetas y toman valores naturales. Pero, aun así, permite

comparar varios modelos, eligiendo aquel que tenga el valor más elevado siempre que se encuentre entre 0 y 1.

3.2.2. Modelo Probit

Tal y como se especificó anteriormente, el modelo Logit y Probit tienen grandes similitudes que han sido explicadas en el apartado anterior, puesto que es el análisis que se va a emplear, por eso, a continuación se explicarán todo aquello que el modelo Probit no comparte con el Logit.

El modelo Probit siguiendo el mismo modelo que en análisis Logit, asume una función de distribución normal (0,1):

$$F(x) = \int_{t=-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

$$Y_i = \int_{t=-\infty}^{\beta_0 + \beta_1 X_{1i} + \beta_k X_{ki}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt + \varepsilon_i \quad i = 1 \dots N$$

Dónde:

$$\int_{t=-\infty}^{\beta_0 + \beta_1 X_{1i} + \beta_k X_{ki}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(X'\beta)$$

Por tanto $Y_i = \Phi(X'\beta) + \varepsilon_i$, cuya función de distribución acumulada normal estándar que se expresa:

$$P(Y_i=1) = P_i = \Phi(X'\beta)$$

$$P(Y_i=0) = P_i = \Phi(-X'\beta)$$

Dónde $\phi(X'\beta)$, es la función de densidad estándar. Ambas distribuciones son similares, la diferencia radica en las colas por lo que los resultados utilizando uno u otro modelo no presentarían grandes cambios, excepto si la muestra fuera grande. La estimación se realiza de la misma manera que con el modelo Logit, teniendo en cuenta los cambios derivados de la función de distribución.

4. BASE DE DATOS

Antes de comenzar a realizar cualquier tipo de análisis para poder llegar a la conclusión de cuáles son las características más influyentes para que un individuo se encuentre o no en situación de desempleo, se debe conocer de dónde proceden los datos que se van a utilizar, así como las distintas modificaciones que se podrían haber realizado para una mayor optimización de los resultados.

La base de datos utilizada es la recogida cada trimestre por el Estado a través de una encuesta, denominada EPA (Encuesta de Población Activa), en la que se recogen los datos de todas las familias que participan en ella, lo que produce un total de 163.262 individuos. También recoge un total de 93 variables en los que se encuentran, entre otras, las variables de la localización

Para un uso más manejable de la base de datos y así obtener unos resultados más simples, se eliminaron todas las variables excepto la situación laboral en el momento de la encuesta, la edad, el sexo, la nacionalidad, el estado civil y el nivel de estudios. El salario, no se ha introducido en el modelo aunque es una variable importante, esto quiere decir que los resultados que se obtengan no serán del todo acertados puesto que el modelo sufre una omisión de variable relevante.

Una vez que la base de datos se queda con las variables seleccionadas, se eliminan todos aquellos individuos denominados inactivos, son aquellos que se encuentran desvinculados del mundo laboral, esta situación puede darse por diversos motivos: ser menor de 16 años (edad mínima legal para poder trabajar), jubilación, incapacidad laboral permanente, etc. Por ello se eliminaron todos aquellos individuos comprendidos entre 0 a 15 años, y mayores de 65 años, que no estaban trabajando ni lo estaban buscando. Tras la eliminación, la muestra de individuos queda en 78.095, es decir, 85.167 personas quedan fuera del estudio.

Una vez que la muestra ha sido filtrada hasta obtener los datos necesarios, se puede empezar a trabajar con ella.

5. VARIABLES UTILIZADAS

Como se ha comentado en el apartado anterior, se han seleccionado una serie de variables para poder realizar el análisis con los datos aportados por la EPA, a continuación se procederá a analizar cada una de ellas, así como la codificación utilizada para poder usar los datos en el programa Eviews mediante el análisis Logit. En total han sido seleccionadas: situación laboral, edad, sexo, nacionalidad, estado civil y nivel de estudios (hasta estudios primarios, 1º estudios secundarios, 2º estudios secundarios y estudios superiores).

El modelo se corresponderá con la ecuación siguiente:

$$\text{SitLab}_i = F (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + \beta_7 X_{7i}) + \varepsilon_i$$
$$i = 1 \dots N$$

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

5.1. Situación Laboral (SitLab_i)

La situación laboral, es la variable a estudiar, es decir, es la variable endógena (Y). Esta variable solo podrá arrojar dos posibles resultados, 0 y 1. Tomará el valor 1 cuando el individuo se encuentre en situación de desempleo y 0, cuando no lo esté.

$$\text{SitLab}_i \begin{cases} 0 & \text{Si el individuo está empleado} \\ 1 & \text{Si el individuo está desempleado} \end{cases}$$

5.2 Edad (X_{1i})

La variable edad, se corresponde a una variable explicativa (X₁), es la única de las variables que no va a ir codificada, pero sí, que irá en intervalos de grupos de edad, desde los 16 años (edad en la que ya se puede acceder al mercado laboral), hasta más de 65 (edad de jubilación). Como se necesita esta variable de manera continua, se ha procedido a escoger el valor más alto del intervalo de edad para poder utilizarla, es decir, en el intervalo “de 16 hasta 19 años” se ha escogido el valor de 19 años para todos los individuos, “de 20 a 24 años” se escoge el valor 24 y se les atribuye a todos, y así con todos los demás, excepto

con el último, que recogía a todos los individuos de “65 años o más” por lo que se ha elegido dejar los 65 años para todos.

Esta variable, al ser la única continua, se utilizará para cruzarla con el resto de variables seleccionadas

5.3. Sexo (X_{2i})

El sexo corresponde a una variable explicativa (X_2). En la EPA, está dividido en dos, hombre y mujer. A la hora de la codificación, se ha procedido a corresponder el valor 0 a las mujeres y el 1 a los hombres

$$\text{Sexo} \begin{cases} 0 & \text{Mujer} \\ 1 & \text{Hombre} \end{cases}$$

5.4. Nacionalidad (X_{3i})

Al igual que en el caso del Sexo, la nacionalidad se ha dividido en dos, aunque en la EPA se divide en tres (español, extranjero, español con doble nacionalidad) en este caso se ha incluido a aquellos individuos con doble nacionalidad como españoles. El valor 0 corresponde a los individuos con nacionalidad extranjera y 1 para los que tienen nacionalidad española (únicamente española o doble nacionalidad)

$$\text{Nacionalidad} \begin{cases} 0 & \text{Extranjero} \\ 1 & \text{Español} \end{cases}$$

5.5. Estado Civil (X_{4i})

Según la EPA, el estado civil contempla cuatro posibles escenarios: soltero/a, casado/a, divorciado/a y viudo/a. En este caso, se han unido las opciones de soltero/a, divorciado/a y viudo/a en uno solo llamado soltero. El valor 0 corresponde a los solteros (soltero/a, viudo/a, divorciado/a) y 1 para los casados. Por lo que queda de la siguiente manera:

$$\text{Estado Civil} \begin{cases} 0 & \text{Soltero} \\ 1 & \text{Casado} \end{cases}$$

5.6. Nivel de estudios

El nivel de estudios es una variable ficticia que se divide en más de una variable explicativa. Para poder abarcar todas las opciones, se ha añadido de manera aditiva al modelo. El nivel de estudios se ha dividido en cuatro: hasta estudios primarios; primera parte de estudios secundarios (abarca de 1º a 3º de la E.S.O.); segunda parte de estudios secundarios (desde 4º de la E.S.O. hasta un grado medio); y estudios superiores (bachillerato, universidad).

Estudios Primarios	$\begin{cases} 0 & \text{Otros} \\ 1 & \text{Hasta E. Primarios} \end{cases}$	2º Secundarios	$\begin{cases} 0 & \text{Otros} \\ 1 & \text{2º secundarios} \end{cases}$
1º secundarios	$\begin{cases} 0 & \text{Otros} \\ 1 & \text{1º secundarios} \end{cases}$	Estudios Superiores	$\begin{cases} 0 & \text{Otros} \\ 1 & \text{E. Superiores} \end{cases}$

Al ser una variable ficticia, no se pueden introducir las cuatro opciones, sino que hay que dejar una de las opciones fuera de la ecuación, en este caso se ha elegido como variable de referencia los 1º estudios secundarios.

6. RESULTADOS

Para comenzar a realizar el análisis, primero se tiene que plantear la ecuación, que constará de siete variables explicativas, la edad, el sexo, la nacionalidad, el estado civil, los estudios primarios, estudios secundarios de 2º orden y estudios superiores. Con los datos obtenidos, podremos asignar un valor y un signo a las betas que serán fundamentales para su posterior análisis. Para este caso, el modelo se especifica de la siguiente manera:

$$\text{Modelo general: } Y_i = \frac{e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + \beta_7 X_{7i}}}{1 + e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + \beta_7 X_{7i}}} + \varepsilon_i$$

Dónde: Y_i = SITLAB (variable dependiente)

X_{1i} = EDAD

X_{5i} = E_PRIMARIOS

X_{2i} = SEXO

X_{6i} = SECUNDARIA2

X_{3i} = NAC

X_{7i} = E_SUPERIORES

X_{4i} = ECIVIL

6.1. Estimación y contrastes

Tras haber especificado el modelo de manera correcta, el siguiente paso es realizar la estimación, se procederá a utilizar el estimador por máxima verosimilitud utilizando un análisis Logit mediante el programa Eviews. Como en cualquier modelo, es necesario realizar distintos contrastes para verificar su validez. Se van a realizar desde la siguiente tabla en la que se encuentran dichos resultados:

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.271625	0.049596	5.476790	0.0000
EDAD	-0.012921	0.000894	-14.45049	0.0000
SEXO	-0.540566	0.019806	-27.29318	0.0000
NAC	-0.334032	0.033071	-10.10043	0.0000
ECIVIL	-0.494836	0.021424	-23.09778	0.0000
E_PRIMARIOS	0.711157	0.034562	20.57612	0.0000
SECUNDARIA2	-0.466529	0.025535	-18.27022	0.0000
E_SUPERIORES	-1.070336	0.024644	-43.43217	0.0000
McFadden R-squared	0.067046	Mean dependent var	0.176247	
S.D. dependent var	0.381033	S.E. of regression	0.368885	
Akaike info criterion	0.869073	Sum squared resid	10625.79	
Schwarz criterion	0.870023	Log likelihood	-33927.14	
Hannan-Quinn criter.	0.869365	Deviance	67854.28	
Restr. deviance	72730.57	Restr. log likelihood	-36365.29	
LR statistic	4876.288	Avg. log likelihood	-0.434434	
Prob(LR statistic)	0.000000			
Obs with Dep=0	64331	Total obs	78095	
Obs with Dep=1	13764			

⁴ Tabla 6.1: Resultados de la ecuación con los datos del INE.

De los coeficientes estimados, se deduce la ecuación del modelo:

$$\hat{Y}_i = \frac{e^{0.27 - 0.01X_{1i} - 0.54X_{2i} - 0.33X_{3i} - 0.49X_{4i} + 0.71X_{5i} - 0.47X_{6i} - 1.07X_{7i}}}{1 + e^{0.27 - 0.01X_{1i} - 0.54X_{2i} - 0.33X_{3i} - 0.49X_{4i} + 0.71X_{5i} - 0.47X_{6i} - 1.07X_{7i}}}$$

De la Tabla 6.1, se pueden extraer distintas conclusiones de las variables explicativas. Por una parte, se puede deducir a través del signo que acompaña a cada uno de los coeficientes, la relación que siguen las variables para posicionarse en una situación de desempleo o de empleo. En caso de que el signo sea negativo, cuanto más aumenta esa variable, menor será la probabilidad de encontrarse desempleado (relación inversa), mientras que si el signo es positivo, cuanto mayor sea el valor, mayor es la probabilidad de encontrarse en paro (relación directa).

⁴ Tabla 6.1: Resultados de la estimación de un modelo logit obtenido en el programa Eviews.

Por otro lado, hay que analizar la significación de las variables, tanto de manera individual como de manera conjunta, para poder así verificar la adecuación de éstas a la hora de explicar el desempleo. Para medir la significación individual las hipótesis que se plantean son:

$$\left\{ \begin{array}{l} H_0: \beta_j = 0 \rightarrow X_j \text{ no será significativa} \\ H_1: \beta_j \neq 0 \rightarrow X_j \text{ será significativa} \end{array} \right.$$

Si la variable no es significativa, significa que la hipótesis nula (H_0) no se rechaza, ocurre lo contrario si la variable si es significativa, la hipótesis nula se rechaza.

Para la significación conjunta, las hipótesis son:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0 \rightarrow \text{si no se rechaza, } X_j \text{ no será significativa}$$

$$H_1: \text{Algún } \beta_j \neq 0 \rightarrow \text{si se rechaza, } X_j \text{ será significativa}$$

A continuación, se exponen los resultados de los contrastes explicados anteriormente:

1. Edad = - 0,012921 \rightarrow La variable edad tiene un valor negativo, por lo que tiene una relación inversa con la variable endógena, a mayor edad, menor es la probabilidad de estar desempleado.

2. Sexo = - 0,540566 \rightarrow La variable sexo tiene también un valor negativo, por lo que el análisis es el mismo, su relación con la variable endógena es inversa, cuando el individuo es hombre, menor es la probabilidad de estar desempleado que la mujer.

3. Nacionalidad = -0,334032 \rightarrow La nacionalidad también mantiene una relación inversa debido a que es un dato con signo negativo. Cuando un individuo es de nacionalidad española tiene menor probabilidad de estar desempleado que la de un extranjero.

4. Estado Civil = -0,494836 \rightarrow El estado civil, al igual que las anteriores, tiene una relación inversa debido a su signo negativo. Por lo tanto, cuando un individuo está casado, menor probabilidad tiene de encontrarse en situación de desempleo que un soltero.

5. Estudios primarios = 0,711157 → Los estudios primarios tienen un signo positivo, por lo que tienen una relación directa con el desempleo. Cuando un individuo tiene un nivel de estudios primarios, mayor es la probabilidad de que se encuentre en paro que uno con estudios secundarios de primera etapa.

6. Estudios 2º secundarios = -0,466529 → Signo negativo, relación inversa con la situación laboral, por lo que cuando un individuo no presenta estudios relacionados con un nivel de 2º secundaria, menos posibilidades tiene de encontrarse en situación de desempleo que uno con estudios secundarios de primera etapa.

7. Estudios superiores = -1,070336 → Signo negativo, por lo que mantiene una relación inversa con la variable endógena, es decir, cuando los estudios del individuo no son superiores, las probabilidades de encontrarse en situación de desempleo aumentan que uno con estudios secundarios de primera etapa.

De manera individual, todas las variables presentan un p.valor = $0 < \alpha = 0,05$, por lo que todas las variables son significativas.

En su conjunto, el modelo presenta una LR = 4876,288 y un p.valor = $0 < \alpha = 0,05$, por lo que el modelo es significativo.

6.2. Bondad del ajuste

En un modelo discreto, como es el análisis logit, las medidas de la bondad del ajuste son significativamente más pequeñas que en otros modelos microeconómicos. La bondad del ajuste, se encuentra entre los valores 0 y 1.

Las dos bondades de ajuste más utilizadas son el Pseudo R^2 de McFadden y el Porcentaje de Predicciones correctas, pero tal y como se especificó anteriormente solo se utilizará el porcentaje de predicciones correctas.

Para utilizar el porcentaje de predicciones correctas se calcula la probabilidad estimada de cada una de las observaciones de la variable endógena (Y), para que tome unos valores determinados.

Se define una tabla de clasificación, en la que se presenta el porcentaje de aciertos frente a los valores observados. De manera habitual, se toma una probabilidad del 0,5 para clasificar a los individuos, en este caso estar

desempleados ($Y_i = 1$) siempre que alcancen una probabilidad (\hat{p}_i) mayor o igual a 0,5. Los resultados se pueden observar a continuación:

Expectation-Prediction Evaluation for Binary Specification
Equation: EQ01
Date: 04/03/20 Time: 11:12
Success cutoff: C = 0.5

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)≤C	64010	13451	77461	64331	13764	78095
P(Dep=1)>C	321	313	634	0	0	0
Total	64331	13764	78095	64331	13764	78095
Correct	64010	313	64323	64331	0	64331
% Correct	99.50	2.27	82.37	100.00	0.00	82.38
% Incorrect	0.50	97.73	17.63	0.00	100.00	17.62
Total Gain*	-0.50	2.27	-0.01			
Percent Gain**	NA	2.27	-0.06			

⁵ Tabla 6.2: Porcentaje de predicciones correctas con los datos del INE.

Del total de las observaciones totales de la muestra (78.095 individuos), el modelo predice correctamente un 82,37% del total, lo que corresponde a un número total de 64.323 observaciones. Teniendo en cuenta estos resultados, se clasificarían correctamente como personas desempleadas a 13.764 (2,27%), mientras que en situación de empleo se posiciona correctamente al 99,5% de las observaciones (64.010 personas). En conclusión se puede afirmar que el modelo predice mejor a los individuos que tienen un empleo, de aquellos que se encuentran en paro. La bondad del ajuste es de 82,37%, es un valor considerado como bueno.

6.3. El efecto marginal de las variables explicativas

El efecto marginal de las variables del modelo, permite la evaluación del efecto que produce cada una de las variables sobre la probabilidad de encontrarse en una situación de desempleo. Este cálculo, se realiza mediante el efecto marginal de las variables explicativas (X_{ji}), que es el efecto que se produce sobre la probabilidad de estar desempleado, un cambio infinitesimal en las variables independientes.

⁵ Tabla 6.2: Porcentaje de predicciones correctas.

En un modelo no lineal, como es el modelo logit, el efecto marginal no se puede deducir con los parámetros que acompañan a los regresores, ni es constante. Se define en la siguiente fórmula, denominada efecto marginal estimado:

$$\frac{\widehat{d\pi}}{dX_{ij}} = dF \frac{X_i' \beta}{dX_{ij}} = f(\widehat{X_i' \beta}) \hat{\beta}_j$$

Identificando f , como la función de densidad de la distribución de una variable logística. Dado que esta función siempre tiene signo positivo, el signo del efecto marginal viene determinado por el signo que determine el coeficiente β de la variable explicativa que se estudie. Al depender de los valores de las X , no es un resultado fijo y por tanto solo se puede conocer o para determinados valores de los regresores o su efecto medio.

Las variables en las que se estudiará dicho efecto, serán solo en aquellas variables explicativas, cuyo carácter sea cuantitativo y sea continua, en este caso, será la variable Edad ya que es la única que es continua y sin encontrarse en el intervalo de 0 y 1 en el que se han codificado el resto de variables cualitativas (sexo, estado civil, nacionalidad y nivel de estudios).

El resto de variables, al ser categóricas se calcula el efecto sobre la probabilidad estimada cuando la variable obtiene el valor 1 y restando la probabilidad estimado cuando vale 0, manteniendo el resto de las variables constantes en la media.

A continuación, se calcularán los efectos en las distintas variables, comenzando por la variable cuantitativa con la que se determinará su efecto sobre la variable independiente (SITUACIÓN LABORAL):

$$\frac{\widehat{d\pi}}{dEDAD_i} = \frac{dFX_i' \beta}{dEDAD_i} = f(\widehat{X_i' \beta}) \hat{\beta}_1 = -0,001709$$

Para calcular el efecto marginal en el programa Eviews, se debe generar la fórmula para obtenerlo, puesto que el programa no dispone de un comando que calcule de manera directa la solución. Para ello, las variables han sido clasificadas en cuantitativas o cualitativas:

➤ **Quantitativas:** El efecto marginal se calcula con el cambio en la probabilidad de estar desempleado ante un pequeño cambio en la edad. El resultado ha sido -0,001709; esto quiere decir, que al aumentar la edad en un año, la probabilidad de encontrarse en situación de desempleo disminuye en un 0,171%. Como se ha dicho anteriormente, la función de densidad es siempre positiva, por lo que es el valor del coeficiente estimado β de la variable EDAD (-0,01292) el que ha proporcionado el signo negativo a este resultado.

➤ **Cualitativas:** Se debe calcular un índice para el modelo cuando toma el valor 1, y otro cuando toma el valor 0. Están en función de los coeficientes estimados para el individuo de referencia y para el valor medio de las demás variables. Para obtener las distintas probabilidades se ha calculado utilizando la fórmula:

$$P_i = \frac{e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + \beta_7 X_{7i}}}{1 + e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + \beta_7 X_{7i}}}; \text{ todas las variables } X$$

se mantienen constantes en la media excepto la variable a estudiar que recibirá el valor 0 o 1. Tras hacer los cálculos, el efecto marginal es el resultado de la diferencia entre la probabilidad de que la variable explicativa tome un valor 1 y cuando toma el valor 0.

Los efectos marginales de cada una de las variables se recogen en la siguiente tabla:

VARIABLE EXPLICATIVA	PROBABLIDAD ESTIMADA (Xi = 1)	PROBABLIDAD ESTIMADA (Xi = 0)	EFECTO MARGINAL
SEXO	0,076657	0,124763	-0,048104
NACIONALIDAD	0,097491	0,132230	-0,034737
ESTADO CIVIL	0,081143	0,127987	-0,046843
E. PRIMARIOS	0,266286	0,151265	0,115019
E. SECUNDARIOS 2	0,100540	0,151265	-0,050725
E. SUPERIORES	0,057592	0,151265	-0,093673

⁶Tabla 6.3. Efectos marginales de las variables cualitativas del modelo. Fuente: Elaboración propia a partir de los datos de Eviews.

⁶Tabla 6.3: Efectos marginales de las variables cualitativas.

La interpretación de las variables anteriores es la diferencia de probabilidad entre que el individuo se encuentre en situación de desempleo cuando presenta la característica representativa y cuando no la presenta. Es decir, la probabilidad de encontrarse en desempleo, disminuye un 9,36% cuando presenta estudios superiores con respecto a los individuos cuyo nivel de estudios se corresponda con los estudios secundarios de primera etapa, mientras que, la probabilidad de encontrarse parado si únicamente posee estudios primarios o inferiores, es un 11,5% mayor frente a otros individuos que tenga un nivel de estudios superiores de primera etapa.

Siguiendo este ejemplo, el análisis del resto de variables es el mismo. Si se estudia la variable sexo, la probabilidad de que los hombres se encuentren en situación de desempleo es un 4,81% menor frente a las mujeres. Atendiendo a la variable nacionalidad, un individuo con nacionalidad extranjera, tendrá una mayor probabilidad de encontrarse en paro frente a otro de nacionalidad española, con una diferencia del 3,47%.

Anteriormente se han explicado los resultados de los estudios primarios y superiores, no dejando sin analizar los individuos que tienen estudios secundarios frente a los que no los tienen, siendo esta diferencia de un 5,07%. Es decir, aquellos individuos que tengan dicho nivel de estudios tendrán menos probabilidades de encontrarse en paro frente a quienes tengan estudios superiores de primera etapa.

6.4. Escenarios

A lo largo de este apartado, se van a analizar las distintas situaciones en las que un individuo con diferentes características de variables, aumentarán la probabilidad de encontrarse en situación de desempleo. Es decir, hasta ahora se han analizado la influencia de las variables según su valor medio, ahora se determinará el valor sobre la probabilidad de ciertas combinaciones de variables.

Se van a plantear distintos escenarios, en los que se enfrentará la edad a cada una de las variables cualitativas (sexo, estado civil, nacionalidad y los distintos niveles de estudios) de manera individual, manteniéndose el resto de las variables cualitativas, constantes. Con ello se determinará la variación en la

probabilidad estimada cuando una característica definida cambia. Los resultados se encuentran en el Anexo I.

A la vista de los resultados, existen escenarios donde las probabilidades de encontrarse en situación de desempleo son muy elevadas en comparación con la tasa de paro actual del 14,5%⁷. El dato más significativo corresponde a una mujer de 24 años, extranjera, soltera y con un nivel de estudios primario, dando un valor del 53,23% de probabilidad de desempleo, siendo un 38,73% mayor que la media de los individuos en edad de trabajar. Si comparamos esta misma situación (mujer de 24 años, soltera y con un nivel de estudios primarios), con nacionalidad española, el resultado sería del 44,9%, que sigue estando muy por encima del porcentaje global de paro, aunque se aprecie una reducción del 8,33%, de lo que se deduce que una mujer de nacionalidad extranjera tiene una probabilidad mayor de encontrarse en una situación de desempleo frente a una mujer con nacionalidad española. Si ahora comparamos ambas situaciones siendo el sujeto un hombre, los porcentajes varían de la siguiente forma: si es extranjero su valor porcentual se eleva al 39,86%, y si es español el porcentaje es del 32,19%. Es decir, que entre el sujeto con mayor porcentaje (mujer, extranjera) y el de menor porcentaje (hombre, español) la diferencia porcentual es del 21,04%.

Si mantenemos las mismas características que en el primer supuesto, modificando la variable edad, de 24 años a 39 años, el porcentaje de desempleo de la mujer extranjera es del 48,39%, y de la mujer española es del 40,17%. Entre ambas, la diferencia es de 8,22%, es una diferencia menor que en el caso anterior cuando las mujeres son más jóvenes. Esto se explica, por la relación inversa entre la edad y el desempleo. Modificando la variable sexo, la probabilidad de un hombre extranjero en desempleo es de un 35,32%, frente a la probabilidad de un hombre español, que es de un 28,11%. En este caso, la diferencia entre el individuo con mayor porcentaje y el de menor porcentaje, es del 20,28%.

⁷ Fuente: Porcentaje del paro en España en el mes de Marzo de 2020, Datos Macro, EXPANSIÓN.
<https://datosmacro.expansion.com/paro/espana?sc=LAB->

Para los individuos que tienen más edad que en los casos anteriormente estudiados (54 años), y presentando las mismas características, las probabilidades obtenidas para cada uno de los cuatro individuos son las siguientes: para las mujeres extranjeras las probabilidades ascienden hasta el 43,58%, y para las mujeres españolas el resultado es del 31,03%, por lo que la diferencia entre ambas es del 12,55%, que en comparación con los anteriores supuestos es mayor. También se observa que a la edad de 54 años predomina la contratación de las mujeres españolas de manera más significativa que en mujeres más jóvenes. Entre los hombres, para los extranjeros la probabilidad es del 31,03%. Mientras que para los españoles es del 24,36%. La diferencia que encontramos entre los hombres es del 6,67%, y entre el porcentaje más alto y el más bajo la diferencia es de 19,22%.

Para la siguiente franja de edad, 65 años, muchos de los individuos ya se encuentran en situación de jubilación, por lo que el número de individuos de la muestra disminuye en comparación con las demás edades. Aun así, se van a realizar los mismos supuestos que para los demás rangos de edad, para las mujeres extranjeras, el porcentaje de desempleo es del 40,12%, para las españolas, es del 32,42%. La diferencia porcentual entre ellas es del 7,7%, es la diferencia más baja que encontramos entre las mujeres de las distintas edades. Para la misma situación pero modificando la variable sexo, los extranjeros tienen un 28,07% frente a los españoles que presentan un 21,84%. La diferencia entre ellos es del 6,23%. La diferencia entre el valor porcentual más alto y más bajo es del 18,28%.

Una vez estudiados todos los supuestos anteriores, la edad es una de las variables que más influye en el valor porcentual, de manera que cuanto mayor es el individuo la probabilidad de encontrarse en una situación de empleo desfavorable disminuye, por lo que su relación es inversa: a mayor edad, menor probabilidad de estar en desempleo.

Para conocer la influencia de la variable “nivel de estudios” sobre la probabilidad de desempleo, se realizarán nuevos análisis, con un nivel de estudios superiores, manteniendo constantes el resto de las variables utilizadas en los supuestos anteriores.

A los 24 años, las mujeres extranjeras con estudios superiores tienen un porcentaje del 16,08%, las españolas tienen una probabilidad 12,27%. Los hombres extranjeros tienen un porcentaje 10,04% y los hombres españoles un porcentaje del 7,4%.

A los 39 años, las mujeres extranjeras tienen una probabilidad del 13,64%, frente al 10,16% que tienen las mujeres españolas. Los hombres extranjeros tienen un porcentaje del 8,42%, y los hombres con nacionalidad española un porcentaje del 6,18%.

A los 54 años, las mujeres extranjeras tienen una probabilidad del 11,51%, y las mujeres españolas una probabilidad del 8,52%. Los hombres extranjeros una probabilidad 7,04% y los hombres españoles un 5,15%.

A los 65 años, las probabilidades son del 10,14% para las mujeres extranjeras, del 7,48% para las mujeres españolas, del 6,17% para los hombres extranjeros y del 4,49% para los hombres con nacionalidad española.

De todo ello se deduce que, al aumentar el nivel de estudios, los porcentajes de desempleo se encuentran todos (excepto mujer extranjera de 24 años) por debajo del porcentaje global de paro, por lo que cuanto mayor sea el nivel de estudios la probabilidad de encontrarse en situación de desempleo es menor, con independencia de sexo y nacionalidad.

Por otra parte, si comparamos los resultados de sexo y nacionalidad, siempre se repite el mismo patrón, las mujeres son más desfavorecidas que los hombres, al igual que los extranjeros frente a los nacionales. Pero, el sexo es un factor más determinante que la nacionalidad, puesto que las mujeres españolas presentan porcentajes de desempleo superiores a los hombres extranjeros. Como conclusión final, se puede deducir que las mujeres extranjeras son las que presentan un mayor nivel de dificultad a la hora de encontrar empleo frente al resto de individuos, independientemente de su nivel de estudios, y que esta última variable es determinante a la hora de conseguir empleo.

6.5. Curvas de respuesta

Para finalizar, se han elaborado una serie de gráficos llamados “curvas de respuesta” que facilitarán la evaluación del efecto que tienen las variables cualitativas sobre la probabilidad de encontrarse en una situación de empleo desfavorable (desempleo o paro) cuando se producen cambios en la variable explicativa “EDAD”.

Para poder construir las curvas de respuesta, se ha procedido a representar de manera gráfica las probabilidades estimadas cuando el individuo presenta una variable explicativa determinada poniéndola en relación a la edad, para obtener como resultado una visión directa de los efectos del desempleo. Cabe destacar que para realizar este apartado se ha tenido que modificar la variable EDAD para poder conseguir datos de forma continua.

En primer lugar, se ha observado que todas las variables presentan grandes diferencias en la probabilidad, por lo que todas son significativas, y por tanto, determinantes a la hora de establecer si un individuo con unas características determinadas tiene más probabilidades de encontrarse en situación de paro. Se van a analizar de manera individual puesto que son variables importantes.

La primera de ellas se corresponde a la curva de edad en función del sexo. Se observa una gran diferencia en la probabilidad de ser un individuo desempleado, siendo las mujeres las más desfavorecidas. Además cuánto mayor es la mujer disminuye la probabilidad de encontrarse en situación de desempleo, por tanto podemos afirmar que las mujeres más jóvenes son las más afectadas a la hora de determinar su probabilidad de estar en el paro frente a los hombres.

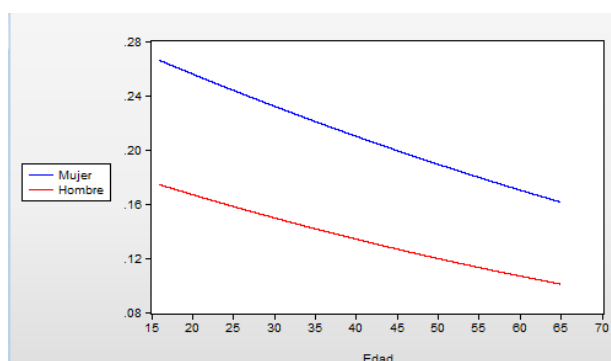


Gráfico 6.1. Curva de respuesta de la edad en función del sexo.

El siguiente gráfico que se va a analizar se corresponde con el estado civil. Los individuos que se encuentran casados son aquellos que menos probabilidades tienen de estar en paro, por lo que estar soltero es una característica que predomina entre los desempleados pero no permanece en el tiempo, puesto que cuanto mayor es la persona soltera, disminuyen sus probabilidades de estar en paro. Por tanto, los jóvenes solteros son los más perjudicados por una situación de empleo desfavorable.

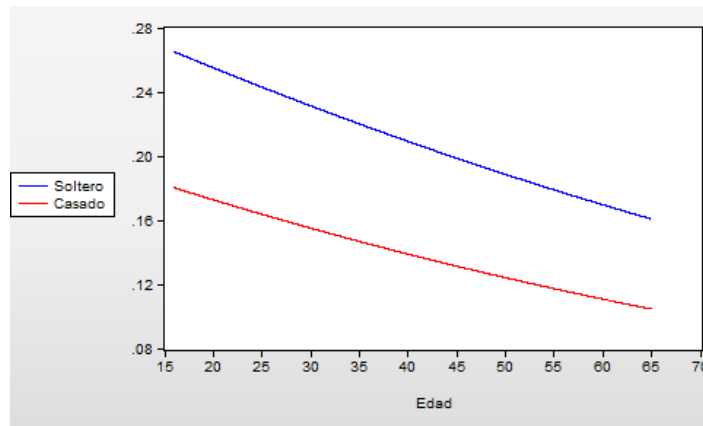


Gráfico 6.2 Curva de respuesta de la edad en función del estado civil.

El siguiente gráfico, se corresponde con la nacionalidad. Se puede observar lo mismo que ha ido ocurriendo en los gráficos anteriores, mucha distancia entre los individuos extranjeros y los españoles que disminuye con la edad aunque las distancias siguen siendo grandes. Los individuos más afectados por esta situación son los jóvenes extranjeros.

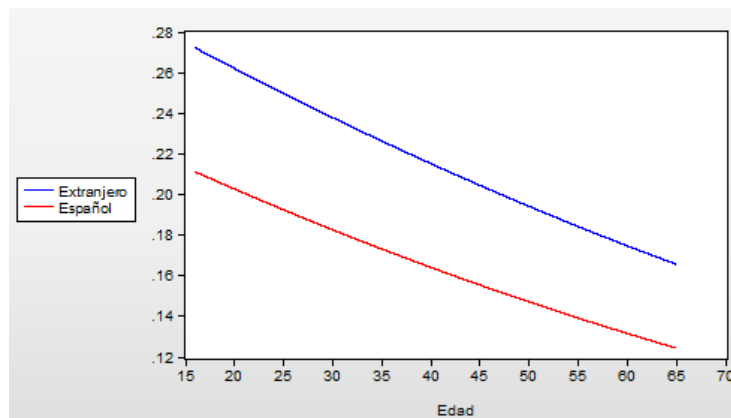


Gráfico 6.3 Curva de respuesta de la edad en función de la nacionalidad.

Por último, se analizará la curva representada por el nivel de estudio que presentan los individuos. La diferencia fundamental con respecto a los gráficos anteriores se debe a que se presentan cuatro variables y no dos. En este caso, aquellas personas cuyos estudios solo llegan hasta la educación primaria tienen una alta probabilidad de encontrarse desempleados. Esta situación disminuye en el tiempo. Mientras que aquellos que han llegado a finalizar sus estudios superiores tienen una probabilidad significativamente inferior que aquellos con niveles de estudios inferiores. Aquellos que presentan un nivel de estudios secundarios de 2º nivel, tienen una probabilidad mucho menor de estar en paro que los que tienen hasta estudios primarios, y, aunque sus probabilidades son similares a los que presentan los superiores, existe una pequeña diferencia entre ellos que se va estrechando a medida que el individuo cumple años. Aunque al igual que el resto de las situaciones, los jóvenes son aquellos que más probabilidades tienen de estar en paro.

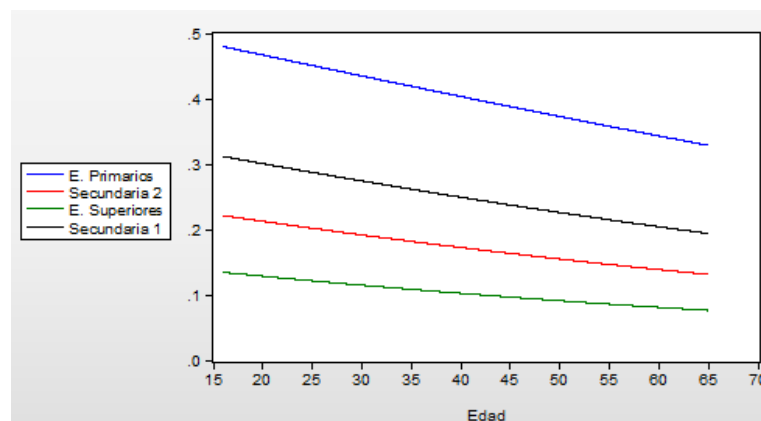


Gráfico 6.4 Curva de respuesta de la edad en función del nivel de estudios

Por tanto, tras todo lo que se ha comentado en ese apartado, se puede concluir que la edad, el sexo, la nacionalidad, el estado civil y el nivel de estudios, son factores determinantes en la situación laboral de los individuos, siendo el nivel de estudios la variable con mayor nivel de influencia en este modelo.

7. CONCLUSIONES

Una vez finalizado el análisis del desempleo en España desde un punto de vista econométrico, se han podido encontrar una serie de características determinantes que influyen directamente en que un individuo se encuentre en esa situación.

Teniendo en cuenta tanto las características como la evolución del desempleo en España, se ha procedido a intentar determinar las características que afectan al desempleo. Para ello, se ha escogido y estimado un modelo Logit para determinar la probabilidad del desempleo en función de unas variables explicativas concretas: la edad, el sexo, la nacionalidad, el estado civil y nivel de estudios de los individuos.

El modelo Logit presentado, predice correctamente el 82,37% de las observaciones, la bondad del ajuste es buena y todas las variables son significativas, tanto a nivel individual como colectivamente, y por todo ello, el modelo es válido. Sin olvidar que no es perfecto puesto que faltan muchas otras variables que harían el modelo mucho más completo, como por ejemplo el salario.

Analizando los resultados, la variable edad muestra una relación inversa: las probabilidades de encontrarse desempleado disminuyen cuanto mayor sea la edad. También disminuye esta probabilidad el hecho de ser hombre frente a ser mujer, ser español frente a ser extranjero, casado a estar soltero, y el nivel de estudios también es influyente, puesto que aquellos con un nivel superior tienen muy pocas probabilidades de encontrarse desempleados frente a los que tienen un nivel de estudios secundarios, estos últimos tienen menos probabilidades que aquellos que solo tienen hasta un nivel de estudios primarios. Con esto se puede deducir que ante cambios en las características, la probabilidad de encontrarse desempleado disminuye a medida que aumenta el valor de las variables.

Para determinar el efecto de las variables dicotómicas se han calculado las probabilidades de que $Y = 0$ y las probabilidades de que $Y = 1$, al obtener el resultado de la diferencia entre ambas probabilidades se obtendrá la probabilidad estar desempleado.

La variable donde mayor peso recae en la influencia de que un individuo se encuentre en una situación desfavorable o no, es el nivel de estudios. Sobre todo cuando el individuo posee un nivel de estudios superiores, lo que disminuye la probabilidad de estar desempleado en un 9,37%.

En los escenarios que se han planteado existen una serie de limitaciones que presenta la probabilidad de una variable mientras las demás permanecen fijas en su media. De ellos se obtienen los cambios en la probabilidad de estar desempleado según la edad, en este caso las probabilidades disminuyen cuando la edad aumenta. A raíz de los resultados obtenidos, los individuos con menor probabilidad de encontrarse en una situación de empleo desfavorable son los hombres españoles, mayores, casados y con estudios superiores, para la situación de la mujer, las características son las mismas, pero tiene mayor probabilidad de estar en paro que un hombre en las mismas condiciones. En caso contrario, aquellos que presentan un mayor porcentaje de probabilidad de estar desempleado, son las mujeres y hombres con un nivel de estudios bajo (hasta estudios primarios), lo que induce a determinar que el nivel de estudios es muy influyente en el momento de obtener un empleo.

No solo se ha realizado un análisis analítico, si no también gráfico mediante las curvas de respuesta. En las gráficas se observa qué nivel de influencia tienen cada una de las variables cuando la variable continua se modifica. Las curvas representan la variación que existe entre los dos valores que puede tomar la variable dicotómica, esa variación es la probabilidad de encontrarse desempleado. Al observar el sexo, no se aprecia una gran diferencia y esta disminuye con la edad, aunque la mujer siempre se encuentra por encima del hombre. Si se analiza el estado civil, las conclusiones son muy similares a las del sexo, a medida que la edad aumenta, las curvas se acercan, siendo los individuos solteros los que más probabilidades tienen de estar desempleados. Lo mismo ocurre en el caso de la nacionalidad, las curvas se unen a medida que el tiempo avanza siendo los más afectados los extranjeros frente a los españoles.

En el gráfico de los distintos niveles de estudios, se puede observar una gran diferencia entre el más bajo (hasta estudios primarios) y el resto, por tanto, la

probabilidad de encontrarse desempleado con un nivel de estudios primarios o menos es notablemente superior al resto de niveles educativos. Pero, al igual que en los escenarios anteriores, las curvas se acercan a medida que la edad es mayor.

Teniendo en cuenta todo lo anterior, **se determina que los individuos con mayor probabilidad de encontrarse en situación de desempleo son las mujeres jóvenes, extranjeras, solteras y con un nivel de estudios primarios o sin estudios.**

En definitiva, haciendo referencia al objetivo fundamental de este Trabajo de Fin de Grado, se puede confirmar que la edad, el sexo, la nacionalidad, el estado civil y el nivel de estudios son factores determinantes en el desempleo.

8. BIBLIOGRAFÍA

“Evolución del mercado de trabajo en España desde 1976”, Diario El País sección Economía, Cinco Días, 28/07/2014 [En Línea], en https://cincodias.elpais.com/cincodias/2014/07/28/graficos/1406544661_293795.html

“Los políticos ya preocupan casi tanto como el paro: Máximo histórico entre problemas que más importan en España”, 20 Minutos, 16/01/2020 [En línea] <https://www.20minutos.es/noticia/4118484/0/indice-datos-preocupacion-politica-espanoles-cis-diciembre-2019/>

https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176918&menu=resultados&secc=1254736195129&idp=1254735976595#!tabs-1254736030639

INE (Instituto Nacional de Estadística). Encuesta de Población Activa (EPA) del 3º trimestre de 2019. Disponible en:

Jiménez, A. (2018): “Así ha evolucionado el paro en España en 2018”, Blog El Blog Salmón, 14/12/2018 [En línea], <https://www.elblogsalmon.com/mundo-laboral/asi-ha-evolucionado-paro-espana-2018>

Porcentaje del paro en España en el mes de Marzo de 2020, Datos Macro, EXPANSIÓN. <https://datosmacro.expansion.com/paro/espana?sc=LAB->

Prieto Alaiz, M. (2019). Análisis Multivariante: Modelos de elección discreta [apuntes de clase].

Wooldridge, J. (2010) Introducción a la Econometría. (4ª ed.) México: Cengage Learning.

Alamilla, N; Arauco, Sigfredo (2009) Limitaciones del modelo lineal de probabilidad y alternativas de modelación microeconómica. [En línea] http://www.utm.mx/edi_anteriores/Temas39/1ENSAYO%2039-1.pdf

ANEXOS

ANEXO I. ESCENARIOS.

Edad	Sexo	Nac	E. Civil	E. Prim.	Secund 2	E. Sup.	\hat{p}
24	0	0	0	1	0	0	53,23%
24	0	1	0	1	0	0	44,90%
24	0	0	1	1	0	0	40,96%
24	1	0	0	1	0	0	39,86%
24	0	0	0	0	0	0	35,85%
24	0	1	1	1	0	0	33,19%
24	1	1	0	1	0	0	32,19%
24	1	0	1	1	0	0	28,78%
24	0	1	0	0	0	0	28,58%
24	0	0	0	0	1	0	25,96%
24	0	0	1	0	0	0	25,42%
24	1	0	0	0	0	0	24,56%
24	1	1	1	1	0	0	22,44%
24	0	1	0	0	1	0	20,06%
24	0	1	1	0	0	0	19,61%
24	1	1	0	0	0	0	18,90%
24	0	0	1	0	1	0	17,61%
24	1	0	0	0	1	0	16,95%
24	1	0	1	0	0	0	16,56%
24	0	0	0	0	0	1	16,08%
24	0	1	1	0	1	0	13,27%
24	1	1	0	0	1	0	12,75%
24	1	1	1	0	0	0	12,44%
24	0	1	0	0	0	1	12,07%
24	1	0	1	0	1	0	11,07%
24	0	0	1	0	0	1	10,46%
24	1	0	0	0	0	1	10,04%
24	1	1	1	0	1	0	8,18%
24	0	1	1	0	0	1	7,72%
24	1	1	0	0	0	1	7,40%
24	1	0	1	0	0	1	6,37%
24	1	1	1	0	0	1	4,65%

Edad	Sexo	Nac	E. Civil	E. Prim.	Secund 2	E. Sup.	\hat{p}
39	0	0	0	1	0	0	48,39%
39	0	1	0	1	0	0	40,17%
39	0	0	1	1	0	0	36,37%
39	1	0	0	1	0	0	35,32%
39	0	0	0	0	0	0	31,53%
39	0	1	1	1	0	0	29,04%
39	1	1	0	1	0	0	28,11%
39	1	0	1	1	0	0	24,98%
39	0	1	0	0	0	0	24,79%
39	0	0	0	0	1	0	22,41%
39	0	0	1	0	0	0	21,92%
39	1	0	0	0	0	0	21,15%
39	1	1	1	1	0	0	19,25%
39	0	1	0	0	1	0	17,13%
39	0	1	1	0	0	0	16,74%
39	1	1	0	0	0	0	16,11%
39	0	0	1	0	1	0	14,97%
39	1	0	0	0	1	0	14,40%
39	1	0	1	0	0	0	14,05%
39	0	0	0	0	0	1	13,64%
39	0	1	1	0	1	0	11,20%
39	1	1	0	0	1	0	10,75%
39	1	1	1	0	0	0	10,48%
39	0	1	0	0	0	1	10,16%
39	1	0	1	0	1	0	9,30%
39	0	0	1	0	0	1	8,78%
39	1	0	0	0	0	1	8,42%
39	1	1	1	0	1	0	6,84%
39	0	1	1	0	0	1	6,45%
39	1	1	0	0	0	1	6,18%
39	1	0	1	0	0	1	5,31%
39	1	1	1	0	0	1	3,86%

Edad	Sexo	Nac	E. Civil	E. Prim.	Secund 2	E. Sup.	\hat{p}
65	0	0	0	1	0	0	40,12%
65	0	1	0	1	0	0	32,42%
65	0	0	1	1	0	0	29,00%
65	1	0	0	1	0	0	28,07%
65	0	0	0	0	0	0	24,76%
65	0	1	1	1	0	0	22,63%
65	1	1	0	1	0	0	21,84%
65	1	0	1	1	0	0	19,22%
65	0	1	0	0	0	0	19,07%
65	0	0	0	0	1	0	17,11%
65	0	0	1	0	0	0	16,71%
65	1	0	0	0	0	0	16,08%
65	1	1	1	1	0	0	14,56%
65	0	1	0	0	1	0	12,87%
65	0	1	1	0	0	0	12,56%
65	1	1	0	0	0	0	12,07%
65	0	0	1	0	1	0	11,18%
65	1	0	0	0	1	0	10,73%
65	1	0	1	0	0	0	10,46%
65	0	0	0	0	0	1	10,14%
65	0	1	1	0	1	0	8,26%
65	1	1	0	0	1	0	7,92%
65	1	1	1	0	0	0	7,72%
65	0	1	0	0	0	1	7,48%
65	1	0	1	0	1	0	6,83%
65	0	0	1	0	0	1	6,44%
65	1	0	0	0	0	1	6,17%
65	1	1	1	0	1	0	4,99%
65	0	1	1	0	0	1	4,69%
65	1	1	0	0	0	1	4,49%
65	1	0	1	0	0	1	3,85%
65	1	1	1	0	0	1	2,79%