



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:

Ana Isabel Heras Portillo

Valladolid, 21 de junio de 2020

RESUMEN

España se sitúa como el segundo país de Europa con mayores tasas de desempleo. Por ello, resulta interesante realizar un análisis de los determinantes que influyen sobre este desempleo.

Como se quiere conocer cómo afectan las características personales del individuo a la probabilidad de que éste se encuentre en situación de desempleo, el modelo elegido es un logit. A través de un modelo logit se puede analizar la influencia de cada determinante a la tasa de desempleo del país.

Tras el análisis de los resultados del estudio desarrollado se obtendrán datos explicativos del desempleo español, donde los determinantes más influyentes resultan ser la edad y el nivel de estudios.

Palabras clave: tasa de desempleo, determinantes, sexo, nivel de estudios, edad, nacionalidad, estado civil distribución, modelo logit.

SUMMARY

Spain ranks as the second country in Europe with the highest unemployment rates. Therefore, it is interesting to carry out an analysis of the determinants that influence this unemployment.

Since we want to know how the personal characteristics of the individual affect the probability that they are unemployed, the chosen model is a logit. The influence of each determinant on the country's unemployment rate can be analyzed through a logit model.

After analyzing the results of the study carried out, explanatory data on Spanish unemployment will be obtained, where the most influential determinants turn out to be age and educational level.

Keywords: unemployment rate, determinants, sex, education level, age, nationality, marital status distribution, logit model.

ÍNDICE

1. INTRODUCCIÓN.....	1
2. OBJETIVOS Y JUSTIFICACIÓN.....	5
3. REVISIÓN DEL ESTADO ACTUAL DEL TEMA.....	6
3.1. Evolución histórica del desempleo en España.....	6
3.2. Análisis en función del sexo.....	9
3.3. Análisis en función de la edad.....	10
3.4. Análisis en función del nivel de estudios.....	11
3.5. Análisis en función de la CC.AA.....	12
4. METODOLOGÍA.....	14
4.1. Definición de las variables.....	14
4.2. Selección de las variables.....	17
4.3. Elección del modelo de regresión.....	19
4.4. Estimación del modelo.....	21
4.5. Creación de escenarios.....	27
4.6. Curvas de respuesta.....	31
5. CONCLUSIONES.....	35
6. BIBLIOGRAFÍA Y WEBGRAFÍA	37

1. INTRODUCCIÓN

Dentro de las ciencias, la **economía** es interpretada de formas muy diversas. A la hora de concretar una definición sobre el término *economía*, no existe una sola que unifique los diferentes planteamientos de los distintos economistas y entendidos de la materia. La definición estándar del concepto economía se traduce a: ciencia que estudia los recursos, la creación de riqueza y la producción, distribución y consumo de bienes y servicios, para satisfacer las necesidades humanas.

Esta ciencia imperfecta provoca inestabilidad por el simple transcurso de la misma. Uno de sus principales desajustes es el desempleo.

Se entiende como **desempleo** la situación en la que se encuentra aquella parte de la población, la cual aun buscando trabajo activamente y estando capacitado para realizar el mismo, no lo encuentra, pasando de tal forma a formar parte del porcentaje de población desocupada del país, es decir, convirtiéndose en población desempleada.

El desempleo se provoca en una situación de desequilibrio entre oferta y demanda, en la que la población que se oferta para trabajar resulta ser mayor que la demanda existente en el mercado por parte de las empresas que hay en él. El porcentaje óptimo se sitúa por debajo al 5% para el desempleo de un país.

La forma de medir el porcentaje de población desempleada es la **tasa de desempleo**, y se calcula como el total de Parados entre el total de Activos.

El **nivel de desempleo** de un país no solo muestra la cantidad de población no ocupada existente, sino que también es indicador de otros posibles problemas estructurales y económicos. Así mismo, el desempleo cuenta con múltiples y variadas causas: puede darse como resultado de un reparto no equitativo de los recursos, un exceso de mano de obra no

cualificada por problemas en la estructura educativa pública, políticas de empleo estatales no acertadas, un índice de actividad económica reducido, o un bienestar y calidad laboral bajos, entre otros.

De esta forma, el desempleo constituye un problema agravado si se tiene en cuenta que no solo se materializa en un problema **económico** (por el no aprovechamiento correcto del factor trabajo), sino que también se traslada al ámbito **social**, convirtiéndose en un factor desestabilizante del equilibrio familiar, dando lugar tanto al malestar social de la población afectada como del resto de la misma por el incremento de los actos delictivos (consecuencia de la situación precaria de ingresos reducidos del primer grupo, dando lugar a que muchos se vean obligados a realizar este tipo de hurtos).

Para el caso concreto de España hay que tener en cuenta que gran parte del porcentaje de afectados forma parte de la población joven, principalmente menores de 25 años. Además, la tasa de desempleo resulta significativamente superior en el caso de las mujeres.

Un gran problema que afecta a la distribución de desempleo en España es la **estacionalidad** y la **temporalidad** de los contratos. Si bien el paro se ve considerablemente reducido en épocas festivas y verano, se debe principalmente a que se trata de un país costero muy desarrollado en el sector terciario del servicio de la restauración, concretamente en el turismo.

Esta contratación masiva unida a la temporalidad reducida que conlleva y a la de los nuevos contratos elaborados actualmente (debido a las vigentes leyes de contratación que nacieron con el objetivo de facilitar la actividad económica a los empresarios), dan lugar a una alta tasa de desempleo y a un empleo poco estable, es decir, que un alto porcentaje de la población que busca trabajo no lo encuentra, y parte de la que lo encuentra es solamente por unos meses o en condiciones laborales no óptimas (jornadas extensas y horarios poco flexibles,

reducida adecuación de los estudios al empleo desempeñado, relación trabajo-salario descompensada, etc.).

Por tanto, teniendo en cuenta las condiciones específicas en las que se encuentra España como unos de los principales países receptores de turismo, resulta comprensible entender la importancia que tiene la estacionalidad en la estructuración de su empleo.

Así mismo las facilidades legislativas vigentes a la hora de extinguir los contratos de los trabajadores también han permitido a los empresarios aprovechar la situación para reducir al mínimo posible el número de contrataciones indefinidas en sus plantillas, reduciendo costes, pero dando lugar a una gran red de empleo temporal.

Uniendo todos estos factores, España se sitúa a la cabeza de los países europeos con mayores tasas de desempleo con un 14,1% a enero de 2019, tan solo por debajo de Grecia (18%). En cuanto al desempleo juvenil anteriormente comentado, la situación resulta aún más complicada, duplicando la tasa general de paro y alcanzando el 32,6% a datos de enero de 2019.

Para el periodo actual, se puede comprobar como la evolución de los resultados con respecto al año anterior 2019 no experimenta apenas variación, pasando de un 14.1% en enero de 2019 a un 13.82% de paro en el 3º trimestre de 2020.

El objetivo de este estudio se basará en la determinación del peso de cada una de las variables que influyen en la probabilidad de estar desempleado del individuo y su importancia, a saber: sexo, nivel de estudios, nacionalidad, estado civil, y edad.

Sin embargo, se tiene en cuenta que las variables que se van a estudiar no comprenden absolutamente todas las que en un entorno real crean influencia sobre las tasas de paro¹, sino las más significativas y accesibles.

De tal forma que se lleguen a alcanzar resultados sobre su incidencia y repercusión en el desempleo español.

Para el desarrollo de este estudio se procederá a definir los objetivos y la justificación de la elección e importancia del tema a analizar. Posteriormente, se procederá a contextualizar el problema del desempleo tanto históricamente como el problema en la actualidad (concretamente para el último trimestre de 2019) en función de las diferentes variables. A continuación, se expondrá la metodología empleada y su desarrollo mediante un modelo logit. Para finalmente explicar las conclusiones acerca de los datos obtenidos tras su elaboración.

¹Teniendo en cuenta la accesibilidad a los datos referentes a las variables objeto de estudio, son varias las que no disponen de la suficiente información que resultaría necesaria para su estudio, como es el caso del **salario**. Del mismo modo existen otro tipo de variables cuyo registro es inexistente debido a su naturaleza, y por lo tanto incontable, como es el caso del **esfuerzo de cada individuo por conseguir empleo activamente**, más allá de estar apuntado a la lista del paro. Por tanto, se tiene en cuenta la limitación en este aspecto del trabajo, resultando igualmente interesante conocer qué variables aportan una mayor influencia al problema del desempleo español.

2. OBJETIVOS Y JUSTIFICACIÓN

Esta investigación tiene como objetivo principal estudiar los determinantes del desempleo español, a través del desarrollo de un modelo probabilístico capaz de analizar el conjunto de características personales que hace que un individuo tenga una mayor o menor probabilidad de encontrarse en situación de desempleo. Como objetivo secundario se pretende averiguar cuál de estos determinantes resulta más relevante: edad, sexo, nacionalidad, nivel de estudios, o estado civil.

La realización de este trabajo se justifica en la importancia del desempleo tanto a nivel estructural (afectando a la economía en su conjunto), como a nivel social como se ha comentado anteriormente. Y es que se trata de un tema de gran relevancia especialmente en individuos jóvenes (concretamente de menos de 25 años donde mayores tasas de desempleo se concentran) coincidiendo con el perfil de estudiante medio. Por tanto, se considera que resulta interesante y necesario conocer los factores que dificultan la obtención de dicho empleo. Por último, al haberse realizado un análisis a través de un modelo de regresión, se pueden obtener conclusiones más fiables y acertadas que con la simple predicción basada en la continuación del ciclo económico actual.

3. REVISIÓN DEL ESTADO ACTUAL DEL TEMA

Para la revisión del estado actual del desempleo en España, se procederá a comentar de forma resumida la evolución histórica del mismo, para después pasar a analizar los datos actuales en función de las variables que los determinan e influyen. Estas se concentrarán en las siguientes: sexo, edad, nivel de estudios, y Comunidad Autónoma de procedencia.

3.1. Evolución histórica del desempleo en España:

Como se ha comentado anteriormente, España se coloca entre los principales países con las tasas de paro más elevadas. A pesar de ello, las tasas de desempleo españolas han sufrido muchas variaciones a lo largo de los años, lo que lleva a que, actualmente, el porcentaje de desempleados en España se encuentra por debajo de las cifras máximas históricas registradas en periodos de crisis y recesión (2012-2014), como se puede ver reflejado en el gráfico 3.1.1 que aparece a continuación:

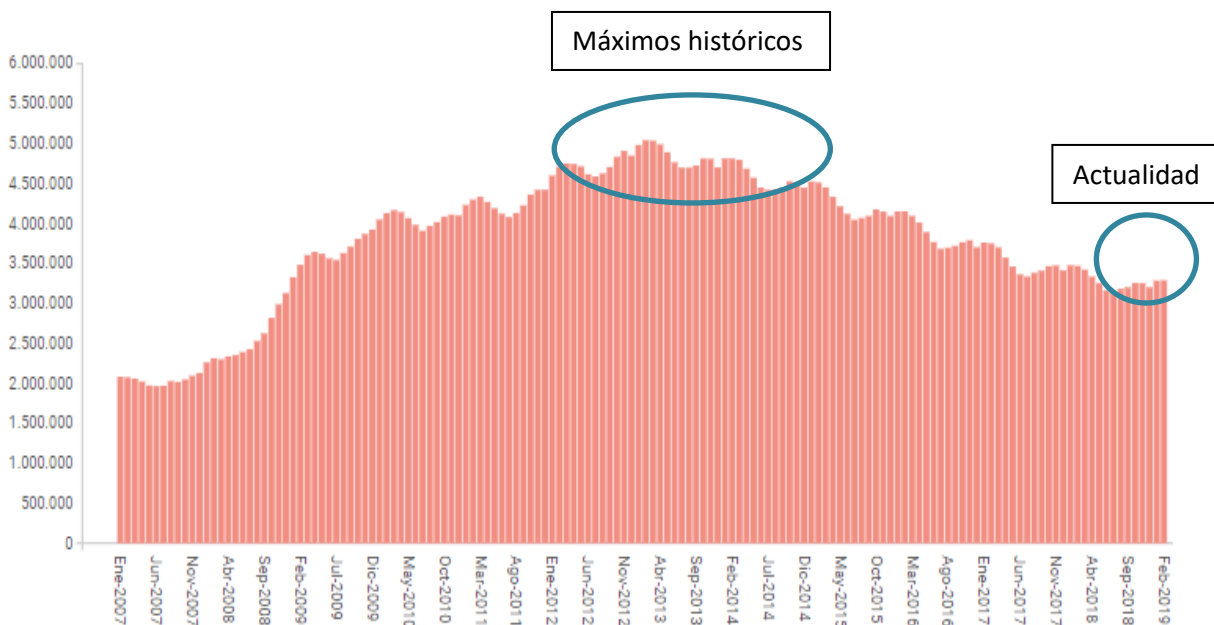


Gráfico 3.1.1.: Evolución del paro registrado desde 2007

Fuente: Ministerio de Empleo y Seguridad Social

En el gráfico 3.1.2 se representa el gráfico de la evolución de la tasa de desempleo española desde 1980, donde se pueden apreciar las diferentes etapas por las que ha pasado su economía:

- Etapa **recesiva** desde finales de los 70 a mediados de los 80
- Etapa **expansiva** desde mediados de los 80 hasta principios de los 90
- Etapa **recesiva** entre 1992 y 1993
- Etapa **expansiva** entre 1994 y 2007
- Etapa **recesiva** desde 2008 hasta 2013
- Etapa **expansiva** desde 2014 hasta la actualidad

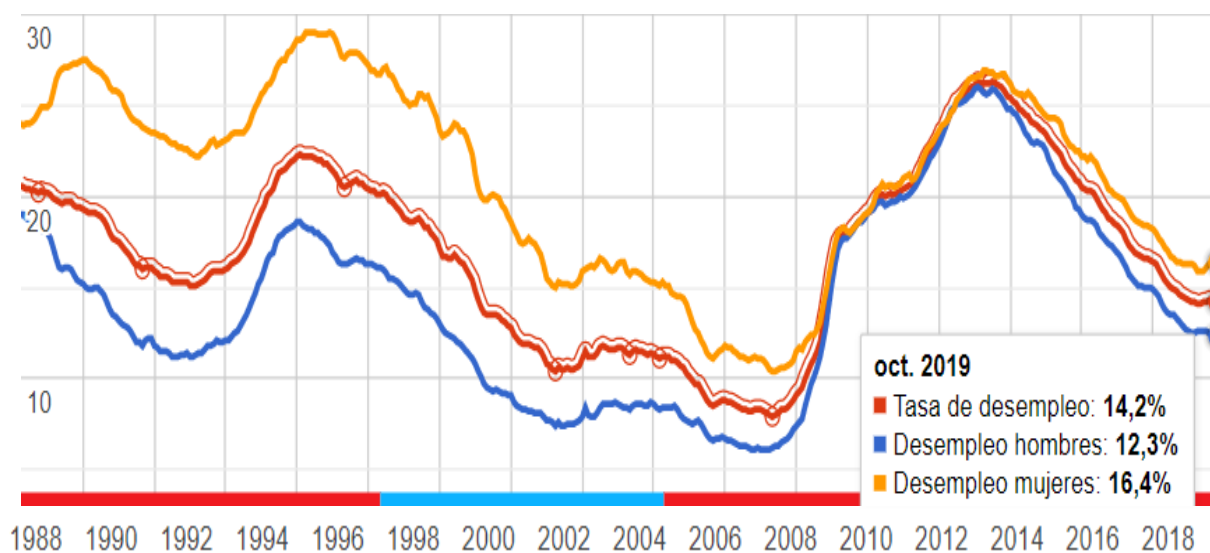


Gráfico 3.1.2: Evolución del paro registrado desde 1980

Fuente: Expansión (Datosmacro.com)

A la hora de analizar la evolución histórica del desempleo en España, resulta de gran importancia tener en cuenta que todos los factores ya comentados influyen en que las fluctuaciones económicas afecten al país por encima de la media. Esto es, que España se encuentra más expuesta a las modificaciones económicas y variaciones cíclicas de la economía en general, ya que cuando la economía global sufre una bajada, la economía española se ve más afectada negativamente que otros países menos sensibles a estos hechos. Del mismo modo, en épocas de crecimiento, la economía española

crece por encima de la media europea. Esto hizo que las tasas de desempleo durante el periodo de crisis económica llegaran a alcanzar máximos históricos.

Esta sensibilidad queda perfectamente reflejada en el siguiente gráfico 3.1.3, en el cual aparece recogida la evolución de la tasa de desempleo en diferentes países del mundo y de la Unión Europea, en comparación con España. En él se puede ver como el comportamiento del paro en España no sigue la tendencia del resto de países europeos, situándose muy por encima estos últimos años, y despuntando en épocas de crisis:

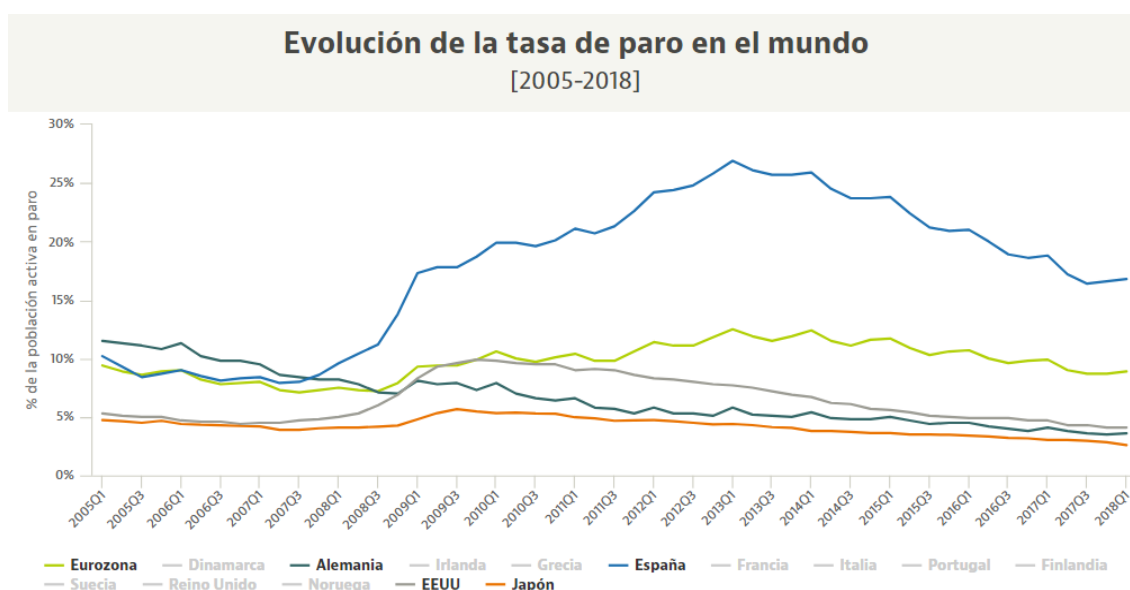


Gráfico 3.1.3: Evolución de la tasa de paro en el mundo desde 2005 hasta 2018

Fuente: INE

Aunque el paro ha disminuido progresivamente con la recuperación económica tras la *superación* de las crisis económicas, en la actualidad la tasa de desempleo en España se sitúa en torno al 14,2% a Octubre de 2019, colocándose en el puesto de segundo país con mayor paro de Europa, por detrás de Grecia. Teniendo en cuenta, también, que ni la economía ni la sociedad de Grecia resulta comparable a la situación española, por lo que resulta de especial interés conocer cuáles son los factores y determinantes que colocan a España en este puesto.

3.2. Análisis en función del sexo:

Como se puede ver representado en el gráfico 3.2, en el que aparece la evolución del desempleo español en función del sexo, la tasa de paro actual (2019T3) se sitúa en el 12,17% en el caso de los hombres, ascendiendo al 15,92% en el caso de las mujeres. Como se puede observar en el gráfico que refleja su evolución, este dato no se trata de un hecho aislado, sino de un comportamiento histórico, en el que la tasa de paro del sector de la población femenino siempre se ha situado con tasas más elevadas que la de los hombres.

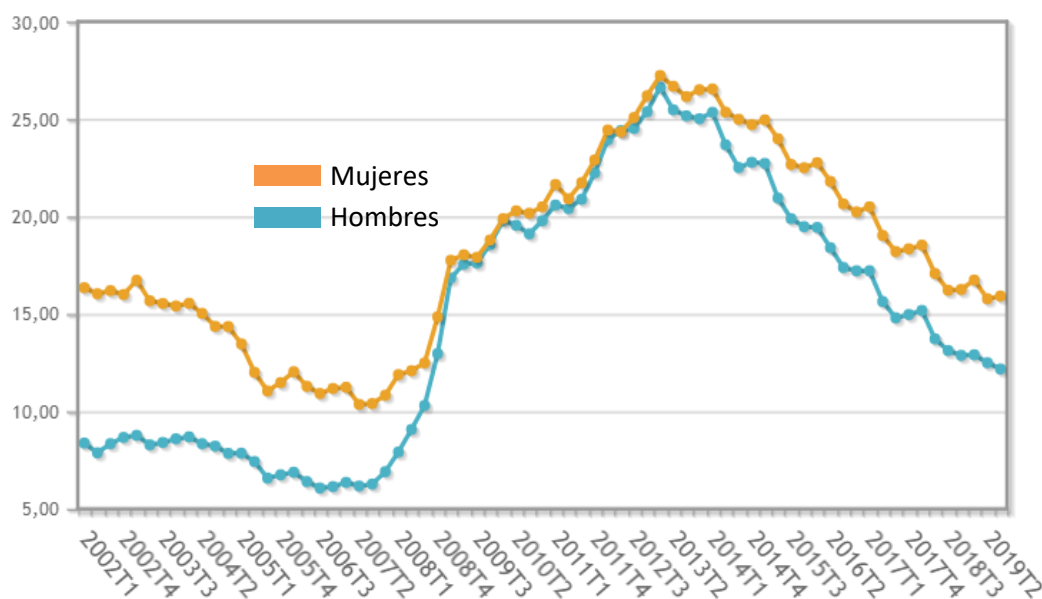


Gráfico 3.2: Evolución del paro registrado desde 2002 en función del sexo

Fuente: Instituto Nacional de Estadística - EPA

De todo ello cabe destacar que, a pesar de no haber vuelto a alcanzar cifras de desempleo tan contrastadas entre ambos sexos como en los años 2000, sí que se puede observar un progresivo incremento de esta desigualdad en los últimos 7 años, desde el *final* de la crisis económica (donde la diferencia entre sexos residía en tan solo 0.1% en 2012) hasta la actualidad (donde está aumentando a casi 4 puntos porcentuales).

Como conclusión del análisis, se puede apreciar que se mantiene una mayor dificultad por parte de la población femenina española a la hora de encontrar trabajo, por encima de la que presentan los hombres. Además, cuando cae el desempleo, las diferencias se agravan.

3.3. Análisis en función de la edad:

En el siguiente gráfico 3.3 se representa la evolución del paro registrado desde 2002 en función de la edad, a través del cual se puede observar cómo la tendencia es fácilmente interpretable. Y es que la tasa de desempleo desciende radicalmente con la edad del individuo. Cabe destacar que, a partir de los 25 años de edad, las cifras de paro prácticamente se equiparan a las de edades más adultas, situándose alrededor de un 12 y 13%. Sin embargo, cuando se trata de la población más joven, estos porcentajes cuadriplican sus cifras, alcanzando en 42,15% de desempleo actualmente, pero habiendo llegado a superar el 75% en 2013.

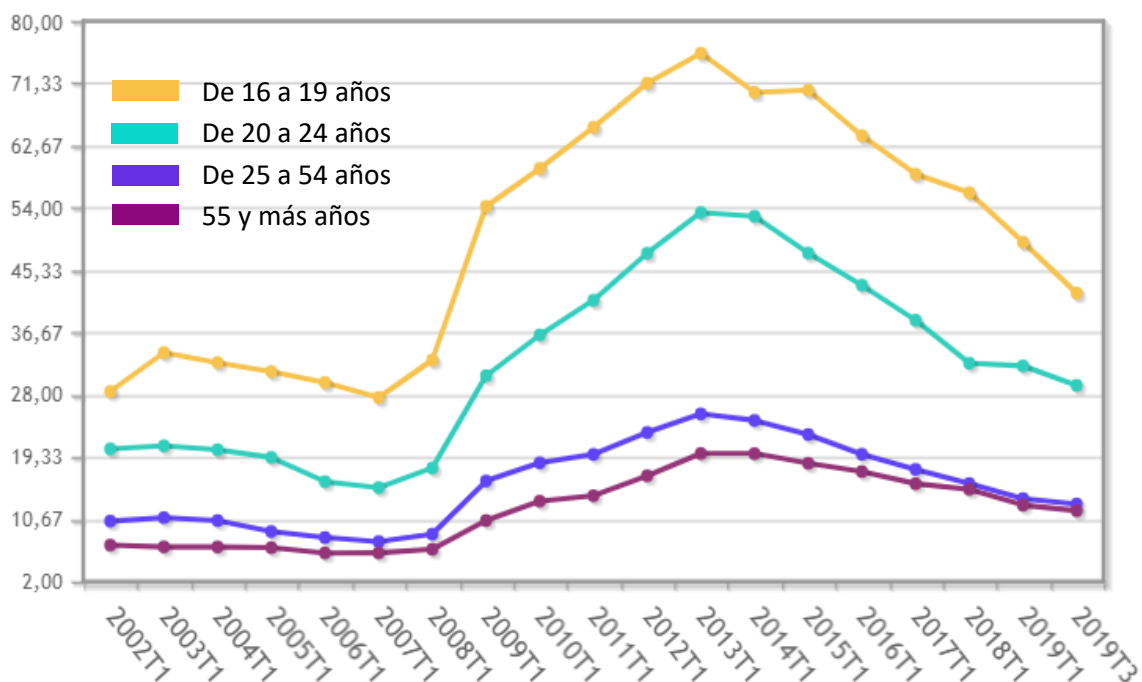


Gráfico 3.3: Evolución del paro registrado desde 2002 en función de la edad

Fuente: Instituto Nacional de Estadística - EPA

En la actualidad se sigue manteniendo esta tendencia, caracterizando el desempleo actual de la siguiente forma: población con 55 o más años (11,79%), población de edades comprendidas entre los 54 y 25 años (12,72%), entre los 24 y los 20 (29,28%), y entre los 19 y los 16 años (42,15%).

Por lo tanto, en el gráfico 3.3 se puede ver reflejada la realidad laboral y social actual, representada a través de esa dificultad añadida de la población más joven para acceder al mercado laboral.

3.4. Análisis en función del nivel de estudios:

El gráfico 3.4 representa la evolución del paro en función del nivel de estudios. Sigue una tendencia directamente proporcional similar a la de la edad, ya que a mayor nivel de estudios la tasa de paro disminuye radicalmente, concentrando el mayor nivel de paro en los núcleos de población analfabeta (36,24%), siguiendo a la población con educación primaria (25,7%), secundaria (14,1%), y por último la población con estudios superiores (8,81%).

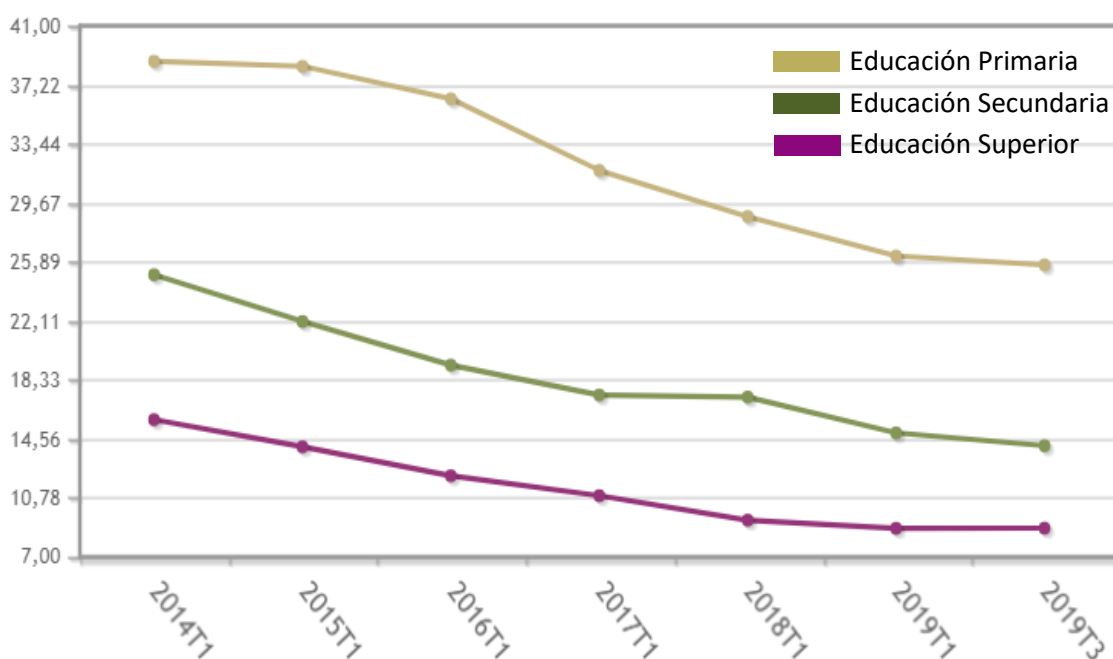


Gráfico 3.4: Evolución del paro registrado desde 2014 en función del nivel de estudios

Fuente: Instituto Nacional de Estadística - EPA

De esta forma se comprueba que efectivamente sigue la tendencia lógica del mercado laboral, donde con mayor nivel de estudios cuenta el individuo, mayor probabilidad tendrá de tener trabajo.

3.5. Análisis en función de la Comunidad Autónoma:

El gráfico 3.5.1 representa la evolución del desempleo en función de la Comunidad Autónoma, a través del cual se puede observar la existencia de diferentes variaciones entre ellas. Destacan Navarra, País Vasco y La Rioja como las Comunidades Autónomas con menor tasa de paro, quedándose en el 8,19%, 9,62%, y 11,11% respectivamente. Mientras que esta tasa se ve considerablemente incrementada en los casos de Melilla, Extremadura, Ceuta, Andalucía y Canarias con un 25,92%, 22,52%, 22,31%, 21,08%, y 21,03% respectivamente.

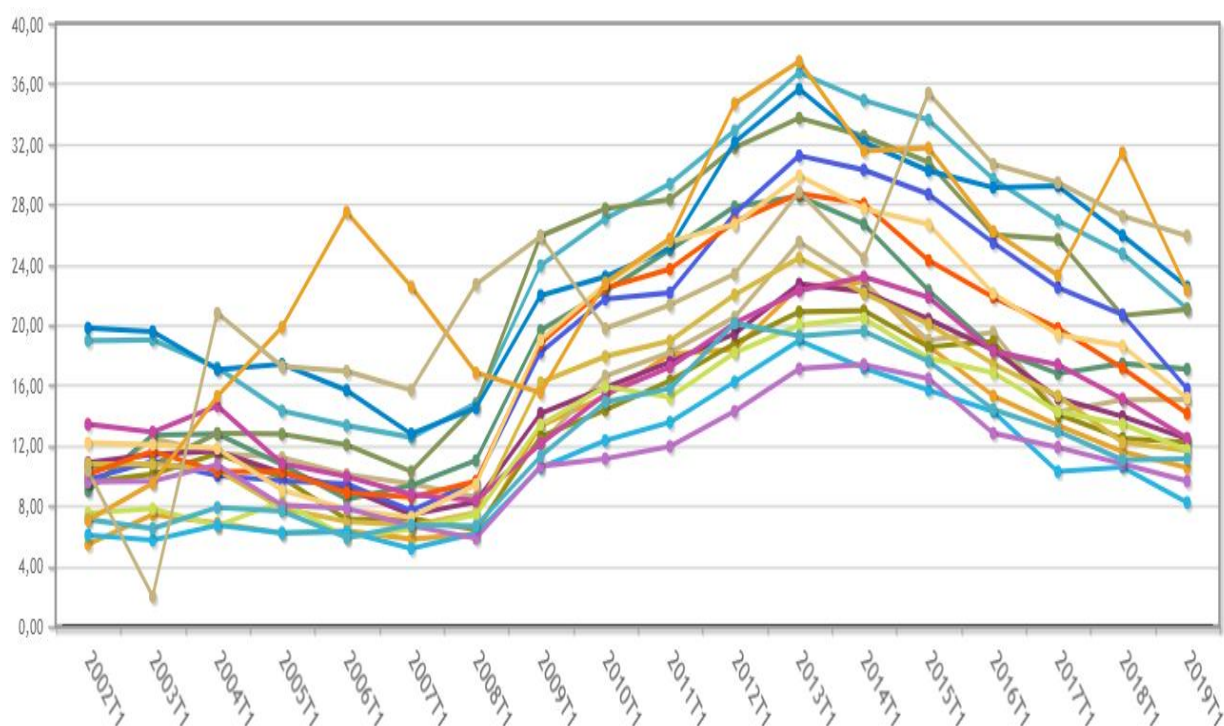


Gráfico 3.5.1. Evolución del paro registrado desde 2002 en función de la Comunidad Autónoma

Fuente: Instituto Nacional de Estadística – EPA

De estos datos se puede concluir la existencia de un cierto patrón Norte-Sur de desempleo, donde las Comunidades de la franja norte experimentan unas tasas menores que las de la sur.

A continuación, se puede observar gráficamente en el siguiente mapa que representa el paro actual en función de la Comunidad Autónoma:

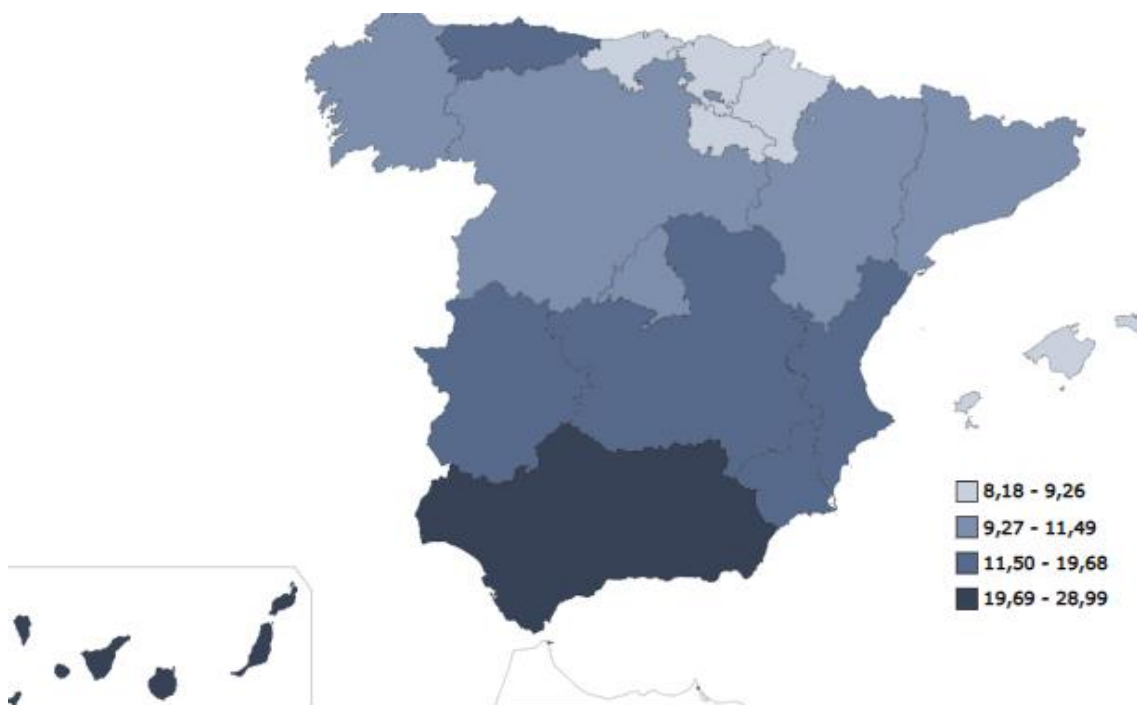


Gráfico 3.5.2.: Mapa del paro actual (2019T3) en función de la Comunidad Autónoma

Fuente: Instituto Nacional de Estadística – EPA

Esto se debe principalmente a que las características tanto de económicas como sociales de cada Comunidad son muy diferentes al tratarse de un país muy plural (distintas flexibilidades de los mercados, estructuras productivas y ciclos económicos), también en mucho más que el ámbito económico (a nivel social, dinámicas de población, etc). Todos estos factores influyen en las fluctuaciones del mercado laboral intrarregional a la hora de valorar el desempleo en España.

4. METODOLOGÍA

4.1. Definición de conceptos:

Las variables que se van a estudiar² se han obtenido de la Encuesta de Población Activa (EPA) del INE.

En primer lugar, se va a proceder a definir los conceptos esenciales empleados según el criterio del INE³:

POBLACIÓN ACTIVA:

La población económicamente activa comprende todas las personas de 16 o más años que durante la semana de referencia tengan empleo, estando disponible para trabajar, y los que se encuentran buscándolo de forma activa⁴.

POBLACIÓN OCUPADA:

Es la formada por todas aquellas personas de 16 o más años que durante la semana de referencia han tenido un trabajo por cuenta ajena o han ejercido una actividad por cuenta propia, según las definiciones dadas a continuación.

A) Son personas con trabajo por cuenta ajena o asalariadas todas las que entren en las siguientes categorías:

- *trabajando*: personas que durante la semana de referencia hayan trabajado, incluso de forma esporádica u ocasional, al menos una hora a cambio de un sueldo, salario u otra forma de retribución conexas, en metálico o en especie.

²Como se ha comentado con anterioridad, las variables a estudiar no recogen completamente todas aquellas que influyen, sino aquellas más significativas de las que se dispone de datos suficientes.

³Todas las definiciones desarrolladas pertenecen a la fuente del Instituto Nacional de Estadística.

⁴La Población Inactiva se daría en el caso de que se dieran las condiciones contrarias, a saber: sin trabajo, ni disponibilidad para empeñarlo, ni en condición de búsqueda activa.

- con *empleo pero sin trabajar*: personas que, habiendo ya trabajado en su empleo actual, estén ausentes del mismo durante la semana de referencia y mantengan un estrecho vínculo con él. La fuerza de este vínculo se determina de acuerdo con la creencia del entrevistado en reincorporarse o no a la empresa y de acuerdo con el periodo de ausencia o la percepción o no de algún tipo de remuneración.

Las personas ausentes de su trabajo por causa de enfermedad o accidente, vacaciones o licencia de maternidad, se consideran como personas con trabajo.

Las ausentes por tener contratos fijos discontinuos, por ser trabajadores estacionales o por estar en espera para incorporarse a un nuevo empleo, se consideran sin empleo.

Los aprendices que hayan recibido una retribución en metálico o en especie y los estudiantes que hayan trabajado a cambio de una remuneración a tiempo completo o parcial se consideran como personas asalariadas y se clasifican como trabajando o sin trabajar sobre la misma base que las demás personas con trabajo por cuenta ajena.

Los miembros activos de las fuerzas armadas figuran también entre la población asalariada.

B) Se considerarán personas con una actividad por cuenta propia todas las incluidas en las siguientes categorías:

- *trabajando*: las personas que durante el período de referencia hayan trabajado, incluso de forma esporádica u ocasional, al menos una hora a cambio de un beneficio o de una ganancia familiar, en metálico o en especie.

- *con trabajo pero sin trabajar*: las personas que durante el período de referencia tenían que realizar algún trabajo a cambio de un beneficio o ganancia familiar pero han estado temporalmente ausentes del mismo por razones de enfermedad o accidente, vacaciones, fiestas, mal tiempo u otras razones análogas.

POBLACIÓN PARADA O DESEMPLEADA:

Según el INE se considerarán paradas todas las personas de 16 o más años que reúnan simultáneamente las siguientes condiciones:

- estar sin trabajo, es decir, que no hayan tenido un empleo por cuenta ajena ni por cuenta propia durante la semana de referencia.
- estar buscando trabajo, es decir, que hayan tomado medidas concretas para buscar un trabajo por cuenta ajena o hayan hecho gestiones para establecerse por su cuenta durante el mes precedente.
- estar disponibles para trabajar, es decir, en condiciones de comenzar a hacerlo en un plazo de dos semanas a partir del domingo de la semana de referencia.

También se consideran paradas las personas de 16 o más años que durante la semana de referencia han estado sin trabajo, disponibles para trabajar y que no buscan empleo porque ya han encontrado uno al que se incorporarán dentro de los tres meses posteriores a dicha semana. Por lo tanto, en este caso no se exige el criterio de búsqueda efectiva de empleo.

Las personas ausentes del trabajo a consecuencia de una suspensión por regulación que no sean consideradas ocupadas, esto es, aquellas cuyo empleador les paga menos del 50 por ciento de su sueldo y que van a reincorporarse a su empleo después de pasados tres meses, son paradas si han buscado trabajo y están disponibles para desempeñarlo.

Si una persona satisface las condiciones de la definición de parado se considerará como tal, aunque compatibilice esta situación con otras como la de estudiante o labores del hogar.

Una vez definidas estas variables, cabe destacar la presencia de la variable bajo la etiqueta AOI, la cual alberga los datos acerca de la situación laboral del individuo, tanto en el caso de ocupación (códigos 03 y 04) como en el de desempleo (códigos 05 y 06). Además, también recoge la parte de la población inactiva (códigos 07, 08, y 09), la cual no corresponde con el objeto del estudio, por lo que la información derivada de los mismos, no será incluida:

AOI = Clasificación de la situación laboral:

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) - *excluida*

08: Inactivos 2 (con los desanimados forman los activos potenciales) - *excluida*

09: Inactivos 3 (resto de inactivos) - *excluida*

4.2. Selección de variables:

Tras la definición de estos conceptos, se va a proceder a la selección de las variables explicativas a estudiar, a saber: sexo, estado civil, edad, nivel de estudios, y nacionalidad. Así mismo, se detallará su composición y los valores asignados. Para ello se ha escogido el periodo para el cual se llevar a cabo el proyecto, obteniendo los datos de la EPA sobre el 3º trimestre de 2019.

- **ACTIVIDAD = AOI:**

- 1: Desempleado
- 0: Ocupado

- **SEXO = SEXO1:** Sexo de la población.

- 1: Hombre
- 0: Mujer

- **ESTADO CIVIL = ECIV1:** Estado civil legal de la población. Solo se pregunta a la población de 16 o más años. En este caso, la variable aparece dividida en 4 condiciones: soltero, casado, viudo, y separado o divorciado. Para facilitar el estudio, éste se limitará a diferenciar entre soltero y el resto de los estados:

- 1: Casado
- 0: Soltero

- **EDAD = EDAD 5.** Edad, grupos quinquenales de años cumplidos de todas las personas. Se trata de una variable continua.

- **ESTUDIOS = NFORMA.** Nivel de estudios de todas las personas de 16 y más años. Esta variable se presenta agrupada en 7 niveles diferentes en función de la formación de los individuos. Para el estudio, se ha simplificado a 3 grupos:
 - Sin estudios⁵
 - Estudios medios⁶
 - Estudios superiores⁷

- **NACIONALIDAD = NAC1.** Nacionalidad de todas las personas. En este caso, la variable también aparece dividida, esta vez en 3 condiciones: nacionalidad española, española y doble nacionalidad, y extranjera. Para facilitar el estudio, se agruparán las 2 primeras, diferenciando entre nacionalidad española o no:
 - 1: nacionalidad española
 - 0: extranjero

Una vez elegidas las variables para el desarrollo del estudio se procederá a concretar tanto la variable dependiente como las variables explicativas:

Y_i = Población desempleada (variable cualitativa)

$$Y_i = \begin{cases} 1 & \text{cuando esté desempleado} \\ 0 & \text{cuando tenga trabajo} \end{cases}$$

⁵En el grupo SIN ESTUDIOS se recogen los datos de los individuos analfabetos, con educación primaria incompleta, y con educación primaria.

⁶En el grupo ESTUDIOS MEDIOS se agrupa la información correspondiente a individuos con estudios de nivel desde la primera etapa de educación secundaria, hasta la segunda etapa de la misma.

⁷En el grupo ESTUDIOS SUPERIORES se recogen los datos tanto de orientación profesional como de educación superior.

Esta variable dependiente estará en función de las siguientes variables explicativas:

$X_1 =$ edad	$X_4 =$ estudios superiores
$X_2 =$ sexo	$X_5 =$ estado civil
$X_3 =$ sin estudios	$X_6 =$ nacionalidad

O bien $X_i =$ conjunto de todas las variables anteriores que influyen en la probabilidad de que una persona esté desempleada.

De esta forma, y tras la elección de un modelo de regresión a continuación, queda un modelo que trata de estudiar la influencia de estas variables sobre el hecho de que una persona se encuentre desempleada.

4.3. Elección del modelo de regresión

A la hora de la elección del modelo de regresión a escoger para la elaboración del análisis se ha optado por el desarrollo de un **modelo logit**. Se ha escogido un modelo logit porque resulta ser el modelo óptimo en comparación con un modelo lineal, en el cual, el hecho de que su estimación es probable que no resida entre los valores 0 y 1, hace que su interpretación resulte imposible, lo que representa el principal problema este modelo.

Otro problema que presenta el modelo lineal es que la varianza es diferente en cada individuo. Esto significa la presencia de heterocedasticidad. Y, es que, teniendo en cuenta un modelo lineal $Y_i = \beta'X_i + u_i$, donde $E(Y_i|X_i) = \beta'X_i$, y su estimación: $\hat{y}_i = \hat{\beta}'X_i$:

su varianza pasa a depender de las X, y por tanto heteroscedastica, ya que las X varían con cada individuo.

$$E(u_i^2) = (1 - \beta'X_i)\beta'X_i = [1 - E(Y_i)]E(Y_i)$$

Sin embargo, en el caso del modelo logit, estos problemas desaparecen.

Un modelo logit es una alternativa para medir la probabilidad de un suceso. Se clasifica dentro de los **modelos de elección discreta**, los cuales tratan de explicar una variable cualitativa en función de otras variables cualitativas y cuantitativas. Dentro de su clasificación, también se incluye dentro de los **modelos de elección binaria**, puesto que se estudian con solo dos alternativas mutuamente excluyentes.

Teniendo en cuenta que el modelo va a tratar de explicar los determinantes que influyen en la situación laboral de un individuo, se tomará como variable dependiente a explicar: el desempleo, la cual tendrá el valor 1 cuando se trate de desempleado, y el valor 0 cuando el individuo tenga trabajo.

La idea del planteamiento del modelo logit consiste en obligar a que las predicciones del modelo se encuentren comprendidas en el intervalo [0,1]. Para lo cual se formula el modelo con una función F que toma sus valores en dicho intervalo, es decir:

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

Donde $0 \leq F(.) \leq 1$ y $E(\varepsilon) = 0$

Por lo tanto:

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

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

En cuanto a la formulación, el modelo logit asume una función de distribución logística de parámetros $\alpha=0$ y $\beta=1$, donde la desviación típica es $\pi/\sqrt{3} = 1,814$

$$F(z) = \frac{\exp(z)}{1 + \exp(z)}$$

$$Y_i = \frac{e^{\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}}} + \varepsilon_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki})}} + \varepsilon_i \quad i = 1 \dots N$$

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

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

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

Para calcular la estimación de máxima verosimilitud se tiene en cuenta que se parte de N observaciones independientes que siguen la siguiente distribución:

$$Y_i = \begin{cases} 1 & p_i = p(Y_i = 1) = F(X_i' \beta) \\ 0 & 1 - p_i = p(Y_i = 0) = 1 - F(X_i' \beta) \end{cases}$$

Donde la densidad de Y_i dado X_i' puede escribirse como:

$$f(Y_i) = [F(X_i' \beta)]^{Y_i} [1 - F(X_i' \beta)]^{1 - Y_i}$$

Por último, su función de verosimilitud es:

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

Y el logaritmo neperiano de la función de verosimilitud:

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

4.4. Estimación del modelo:

Definidas las variables explicativas que van a influir sobre la variable endógena, y el modelo de regresión óptimo para su análisis, se va a proceder a la estimación del modelo de estudio:

En primer lugar, se define el modelo en función de sus variables:

$$Y_i = \beta_0 + \beta_1 \cdot \text{SEXO}_i + \beta_2 \cdot \text{EDAD}_i + \beta_3 \cdot \text{SIN ESTUDIOS}_i + \beta_4 \cdot \text{ESTUDIOS SUPERIORES}_i + \beta_5 \cdot \text{ESTADO CIVIL}_i + \beta_6 \cdot \text{NACIONALIDAD}_i + \omega$$

A través del programa SPSS, se recodifican las variables de forma que se introduzcan ficticias en el modelo para abarcar las categorías del nivel de estudios. Hay que tener en cuenta que el número de ficticias creado debe ser uno menor al número de variables, de tal forma que no se caiga en la trampa de las ficticias. Además, se ha procedido a la eliminación de aquellos individuos cuyos datos eran incompletos.

Una vez recodificadas las variables, se introduce el modelo para su estimación en el programa EViews :

Dependent Variable: DESEMPLEO				
Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)				
Date: 02/24/20 Time: 01:05				
Sample: 1 74864				
Included observations: 74862				
Convergence achieved after 6 iterations				
Coefficient covariance computed using observed Hessian				
Variable	Coefficient	Std. Error	z-Statistic	Prob.
EDAD	-0.024660	0.001003	-24.58316	0.0000
ESTADO_CIVIL	-0.524817	0.024569	-21.36091	0.0000
SEXO	-0.432900	0.022200	-19.50013	0.0000
SIN_ESTUDIOS	0.849884	0.038337	22.16887	0.0000
ESTUDIOSUPERIORES	-0.706839	0.023589	-29.96479	0.0000
NACIONALIDAD	-0.239017	0.036903	-6.476938	0.0000
C	0.083654	0.049954	1.674610	0.0940
McFadden R-squared	0.063132	Mean dependent var	0.136611	
S.D. dependent var	0.343439	S.E. of regression	0.334191	
Akaike info criterion	0.747366	Sum squared resid	8360.073	
Schwarz criterion	0.748228	Log likelihood	-27967.64	
Hannan-Quinn criter.	0.747631	Deviance	55935.28	
Restr. deviance	59704.56	Restr. log likelihood	-29852.28	
LR statistic	3769.275	Avg. log likelihood	-0.373589	
Prob(LR statistic)	0.000000			
Obs with Dep=0	64635	Total obs	74862	
Obs with Dep=1	10227			

Tabla 4.4.1: Estimación del modelo logit.

Fuente: Elaboración propia a través de EViews a partir de los datos de la encuesta sobre el 3º trimestre de 2019 del Instituto Nacional de Estadística – EPA

En lo referente a la significación de las variables, a través del valor de su p valor menor a 0,05, y como puede observarse en la tabla 4.4.1, todas ellas resultan relevantes a la hora de explicar el comportamiento de la dependiente, es decir, todas las variables explicativas que se han añadido al modelo sí que influyen sobre la probabilidad de desempleo. Además, también resultan significativas conjuntamente.

A la hora de interpretar los resultados obtenidos en la estimación, hay que tener en cuenta que los coeficientes de las variables no son interpretables numéricamente, es decir, que el valor proporcionado por la estimación del modelo no refleja la variación real que se produciría como consecuencia de la influencia de la variable explicativa sobre la dependiente. Por lo tanto, solo se puede interpretar su signo.

Y es que $f(-X_i'\beta)$, al tratarse de una variable de densidad, siempre va a ser mayor que cero, siendo además no constante, y diferente para cada individuo. Por lo que solo se puede interpretar el signo de β_j .

Así, el efecto que un incremento de X_i produce en la probabilidad del individuo de encontrarse en situación de desempleo se recogería como:

$$\frac{\partial E(Y_i|X_i, \beta)}{\partial X_{ij}} = f(-X_i'\beta)\beta_j$$

Y para el caso concreto del modelo logit:

$$\frac{\partial E(Y_i|X_i, \beta)}{\partial X_{ij}} = \frac{e^{X_i'\beta}}{(1 + e^{X_i'\beta})} \beta_j$$

De esta forma, se puede observar como el signo negativo de casi todas las variables indica que se tratan de factores que al tomar el valor 1 se reduce la probabilidad del individuo de encontrarse en situación de desempleo.

En la situación inversa se encuentra la variable SIN_ESTUDIOS, la cual cuenta con un signo positivo en el coeficiente de su β . Así, los individuos que posean

un nivel de estudios correspondiente a la primera etapa de educación secundaria o inferior cuentan con un mayor porcentaje de probabilidad de encontrarse en situación de desempleo que aquellos que tienen estudios medios.

Por otro lado, se puede observar como la variable sexo cuenta con un valor negativo, lo cual, recordando que se trata de una variable asignada al valor 1 cuando se trata de hombre, y al 0 en el caso de mujer, da lugar a la interpretación de que los individuos hombres cuentan con menos posibilidades de estar en paro, situando a la mujer en una posición de desventaja.

Este hecho se ve repetido en el caso de la variable NACIONALIDAD, donde, a través de su signo negativo, los individuos extranjeros también cuentan con menos posibilidades de encontrar empleo frente a un español. De los resultados también se deduce que tanto la EDAD como el NIVEL DE ESTUDIOS se comportan de la misma manera, reduciendo las probabilidades de que el individuo esté en paro cuanto mayor nivel de estudios y edad tenga. Lo mismo pasa con el ESTADO CIVIL, donde llama especialmente la atención que la población casada cuenta con mayores probabilidades de formar parte del porcentaje de población ocupada.

Una vez conocido el signo del efecto de cada variable sobre la endógena, resulta interesante conocer la variación numérica concreta que supone cada variable dicotómica sobre la probabilidad final. Para ello, y teniendo en cuenta, como se ha comentado anteriormente, que los coeficientes de las mismas no suponen indicadores reales, se puede recrear cada situación sustituyendo cada variable bien por el valor 1 o 0 y midiendo el efecto que esto tiene en la dependiente. Todo ello manteniendo el resto de variables constantes respecto a su media para aislar el efecto de cada una.

	EDAD	ESTADO_CVIL	SEXO	SIN_ESTUDIOS	ESTUDIOSUPERIORES	NACIONALIDAD
Mean	41.74006	0.546739	0.520758	0.058615	0.517138	0.925770

Tabla 4.4.2: Media de las variables explicativas.

Fuente: Elaboración propia a través de EViews a partir de los datos de la encuesta sobre el 3º trimestre de 2019 del Instituto Nacional de Estadística – EPA

VAR.	PROBABILIDADES	EFFECTO FINAL SOBRE LA P(Desempleo)
ESTADO CIVIL	P(Desempleo / E.Civil = 0) = 0.318119 P(Desempleo / E.Civil = 1) = 0.216319	- 0.101799 (- 10.17%)
SEXO	P(Desempleo / Sexo= 0) = 0.304930 P(Desempleo / Sexo = 1) = 0.221520	- 0.083410 (- 8.34 %)
SIN ESTUDIOS	P(Desempleo / Sin.E. = 0) = 0.3244 P(Desempleo / Sin.E. = 1) = 0.5290	0.2046 (20.46 %)
ESTUDIOS MEDIOS	P(Desempleo / Sin.E. = 0) = 0.2593 P(Desempleo / Sin.E.= 1) = 0.3244	- 0.0651 (- 6.51 %)
ESTUDIOS SUPERIORES	P(Desempleo / E.Sup = 0) = 0.3244 P(Desempleo / E.Sup = 1) = 0.1914	- 0.133 (- 13.3%)
NACIONALIDAD	P(Desempleo / E.Nac= 0) = 0.304049 P(Desempleo / E.Nac = 1) = 0.255953	- 0.048096 (- 4.809 %)

Tabla 4.4.3: Probabilidad estimada en función del efecto de cada variable dicotómica.

Fuente: Elaboración propia a partir de la estimación del modelo logit desarrollado en EViews.

De este modo, en la anterior tabla 4.4.3. quedan reflejadas las probabilidades estimadas. En ella se puede observar el efecto que tiene cada variable sobre la probabilidad de desempleo del individuo.

Empezando por el **ESTADO CIVIL**, el efecto final que tiene sobre la probabilidad de desempleo es negativo. Esto indica que aquellos individuos que se encuentren en situación de matrimonio cuentan con menos posibilidades de formar parte del porcentaje de población desempleada, en contraposición con los individuos solteros, los cuales cuentan con una probabilidad mayor de encontrarse sin trabajo, concretamente un 10,179%.

Siguiendo con la variable **SEXO**, el efecto final también es negativo: - 8,34%, es decir, que los individuos hombres tienen un 8,34% menos de probabilidad de estar desempleados, lo que da lugar a la existencia de diferencias laborales entre hombres y mujeres, teniendo como única variable explicativa del efecto diferenciador el sexo, repercutiendo negativamente en la población femenina frente a la masculina.

En cuanto a los **ESTUDIOS**, el efecto que éstos tienen sobre la probabilidad de desempleo del individuo continúa con la tendencia de la interpretación de los signos de los coeficientes beta de la estimación anteriormente realizada, contando con un 13,3% menos de probabilidad de estar parado aquellos individuos que cuentan con estudios superiores en comparación con aquellos que cuentan con estudios medios. En el lado contrario, se produce un aumento del 20,46% de la probabilidad de desempleo en el caso de los individuos sin estudios frente a con estudios medios. Estos datos presentan una distancia total del 33,76%, lo cual se traduce en un efecto diferencial sobre la probabilidad de estar desempleado del 33,76%, es decir, que la diferencia para un individuo entre no tener estudios y tener estudios superiores, repercute en un 33,76% sobre su probabilidad final de tener empleo. Además, aquellos individuos que cuentan con estudios medios cuentan con un 6,51% menos de probabilidades de encontrarse en paro. Estos datos consolidan la idea de que a mayor nivel de estudios mayor probabilidad de tener trabajo.

Por último, la **NACIONALIDAD** también cuenta con un signo negativo, lo cual indica que los individuos nacionales cuentan con un 4,809% más de probabilidad de encontrarse en situación de actividad, frente a la población extranjera, que verá su posibilidad de estar empleada disminuida en ese porcentaje frente a la población nacional.

4.5. Creación de escenarios:

Otra forma de generar e interpretar el efecto diferencial de cada variable empleada en el modelo son los escenarios. A través de ellos se puede ver con mayor claridad los cambios que se producen en la probabilidad de que un individuo se encuentre en situación de desempleo al realizar un cambio en una

variable del mismo. Además, resultan una forma más precisa de medir el efecto, ya que no se basan en la media de las variables como en el estudio anterior, sino en sus valores concretos.

Para su elaboración se han llevado a cabo diferentes series, contrastando sus resultados teniendo en cuenta ese cambio de característica en cada una.

Los resultados aparecen recogidos en la *tabla 4.5.1* que aparece a continuación. Para su estudio se han dividido los resultados por grupos de edad, y dentro de ellos confrontando una característica frente a la ausencia de la misma.

Nacionalidad

De este modo, y empezando por los individuos más jóvenes, se puede ver la diferencia existente entre dos individuos de 20 años jóvenes, hombres, solteros, y con estudios superiores, tan solo diferentes en cuanto a la característica **nacionalidad**. Se observa cómo existe una diferencia del 3,19% entre ambos, siendo el individuo extranjero el que cuenta con ese porcentaje a mayores por encima de la probabilidad de encontrarse en situación de desempleo frente al individuo de nacionalidad española.

Planteando el mismo escenario (jóvenes, hombres, solteros, y con estudios superiores) pero en individuos de 35 años, se obtienen un resultado similar, en el que el individuo español cuenta con menos posibilidades de estar desempleado. Sin embargo, para esta edad, la diferencia es menor, residiendo en un 2,44% frente al 3,19% anterior. Además, ambas probabilidades de desempleo también resultan significativamente menores en comparación con los anteriores individuos con los que compartían las mismas características a excepción de la edad: 10,35% y 12,79% (con 35 años) frente al 14,32% y 17,51% (con 20 años).

Este mismo hecho se repite en los mismos individuos con 60 años, en los cuales la probabilidad de estar desempleado se reduce aún más en ambos casos y reduciendo al mismo tiempo la diferencia entre ambos, contando con una probabilidad de desempleo del 5,87% en caso del individuo nacional, y de un 7,33% en el del extranjero, dando lugar a una diferencia de tan solo en 1,46%.

EDAD	E. CIVIL	SEXO	E. SUP	SINE	NAC	Ā p
20	NO	H	SÍ	NO	ESP	0,143281
20	NO	H	SÍ	NO	EXTRAN	0,175189
20	NO	H	NO	SÍ	ESP	0,442361
20	NO	H	NO	NO	ESP	0,253228
20	NO	M	NO	SÍ	ESP	0,550160
20	NO	M	SÍ	NO	ESP	0,204988
20	SÍ	M	SÍ	NO	ESP	0,132363
35	NO	H	SÍ	NO	ESP	0,103567
35	NO	H	SÍ	NO	EXTRAN	0,127952
35	NO	H	NO	SÍ	ESP	0,354004
35	NO	H	NO	NO	ESP	0,189791
35	NO	M	NO	SÍ	ESP	0,457953
35	NO	M	SÍ	NO	ESP	0,151189
35	SÍ	M	SÍ	NO	ESP	0,095339
60	NO	H	SÍ	NO	ESP	0,058706
60	NO	H	SÍ	NO	EXTRAN	0,073393
60	NO	H	NO	SÍ	ESP	0,228290
60	NO	H	NO	NO	ESP	0,112259
60	NO	M	NO	SÍ	ESP	0,313224
60	NO	M	SÍ	NO	ESP	0,087719
60	SÍ	M	SÍ	NO	ESP	0,053828
60	SÍ	H	SÍ	NO	ESP	0,035587

Tabla 4.5.1: Escenarios.

Fuente: Elaboración propia a partir de la estimación del modelo logit desarrollado en EViews.

Nivel de estudios

Siguiendo esta comparativa, esta vez variando la característica **nivel de estudios**, el mismo individuo anterior (20 años, hombre, soltero, y español) cuenta con una probabilidad de estar desempleado del 44,23% cuando no tiene estudios frente al mismo individuo con estudios medios, que disminuye al 25,32%, y al 14,32% cuando tiene estudios superiores frente a medios. Por lo tanto, existe una diferencia del 29,91% entre no tener estudios y tener estudios superiores, constituyendo la característica nivel de estudios cómo una de las más relevantes a la hora de explicar su influencia sobre la probabilidad de estar desempleado.

Comparando este mismo escenario pero con individuos de 35 años, estas probabilidades bajan a 35,4%, 18,97%, y 10,35%, respectivamente.

Y en el caso de los individuos con 60 años, bajan aún más hasta 22,82%, 11,22%, y 5,87%, respectivamente.

Por lo tanto, además de observar que efectivamente la probabilidad de estar desempleado aumenta cuanto menor nivel de estudios tiene el individuo, también disminuye conforme aumenta la edad del mismo.

Estado civil

La característica **estado civil** afecta negativamente a los individuos solteros, los cuales cuentan con una mayor probabilidad de estar desempleados (20,49%) frente a aquellos que están casados (13,23%), en el caso de la comparativa entre dos individuos mujeres de 20 años, españolas, y con estudios superiores. Sigue el mismo comportamiento que las anteriores variables, reduciendo la probabilidad de desempleo del individuo a mayor edad del mismo.

Sexo

Variando la característica **sexo** de dos individuos de 20 años solteros, sin estudios, y españoles, se comprueba cómo existe una diferencia del 10,77% entre ambos únicamente por sexo, contando con un 44,23% de probabilidad de estar desempleado en el caso del hombre, frente a un 55,01% para la mujer.

Este hecho se repite independientemente de la edad del individuo recogida en los escenarios, manteniendo esa brecha de entorno al 10% a lo largo de todas las edades del mismo. En la *tabla 4.5.1.* se puede comprobar cómo para individuos de las mismas características (solteros, sin estudios, y españoles) pero de 60 años, este porcentaje se reduce tan solo un 1,5% respecto al 10% existente de diferencia entre los individuos mujer y hombre de 20 años. Esta pequeña variación (1,5%) resulta interesante al compararla con las variaciones que experimentaban las probabilidades anteriores cuando era una de las características anteriormente mencionadas, en lugar del sexo, la que cambiaba en el individuo, y es que, en los casos anteriores se ha podido comprobar la existencia de un descenso significativo de la probabilidad de desempleo a mayor edad del individuo, mientras que en el caso de la brecha existente por sexos, se puede ver cómo se mantiene a lo largo de las distintas edades sin experimentar un descenso significativo.

Por lo tanto, se puede decir que existe una brecha laboral estructural de género, en perjuicio de las mujeres, las cuales en posición de tener las mismas características que un hombre, cuentan con más probabilidades de encontrarse en situación de desempleo.

Para finalizar con los comentarios sobre las probabilidades de desempleo obtenidas en la realización de los escenarios, cabe destacar la existencia de una misma tendencia respecto a la edad, y es que, a mayor edad del individuo, menor probabilidad de encontrarse en situación de desempleo tiene el mismo, siendo los individuos de 60 años, en este caso, los que presentan unas probabilidades más reducidas, con diferencias de hasta el 20% en comparación con individuos de iguales características pero de menor edad.

Además, llama la atención que la mayor probabilidad de encontrarse en desempleo llega al 55%, perteneciente a una mujer de 20 años, soltera, sin estudios, y extranjera. En el lado contrario, se encuentra un hombre de 60 años, casado, con estudios superiores, y español, contando tan solo con el 3'55% de probabilidades de no tener empleo.

4.6. Curvas de respuesta:

Antes de poder proceder a la realización de las curvas de respuesta, es necesario contar con una variable continua sobre la que medir el nivel de desempleo. En este caso dicha variable será la edad. Sin embargo, los datos proporcionados por la EPA disponen la variable en intervalos, a saber: 16-20-25-30-35-40-45-50-55-60-65. Por ello, primero se convertirá la variable edad a una variable continua, para después poder proceder a la elaboración de las curvas de respuesta en función de la edad, a través de la creación de diferentes series en las que se irá variando cada característica por 0 o 1, manteniendo el resto de variables en su media, y observando las diferencias existentes entre las series creadas. Para ello, se llevarán a cabo diferentes escenarios, y se confrontarán, de tal forma que se pueda observar gráficamente la diferencia existente a la aportación sobre la probabilidad de desempleo de cada variable.

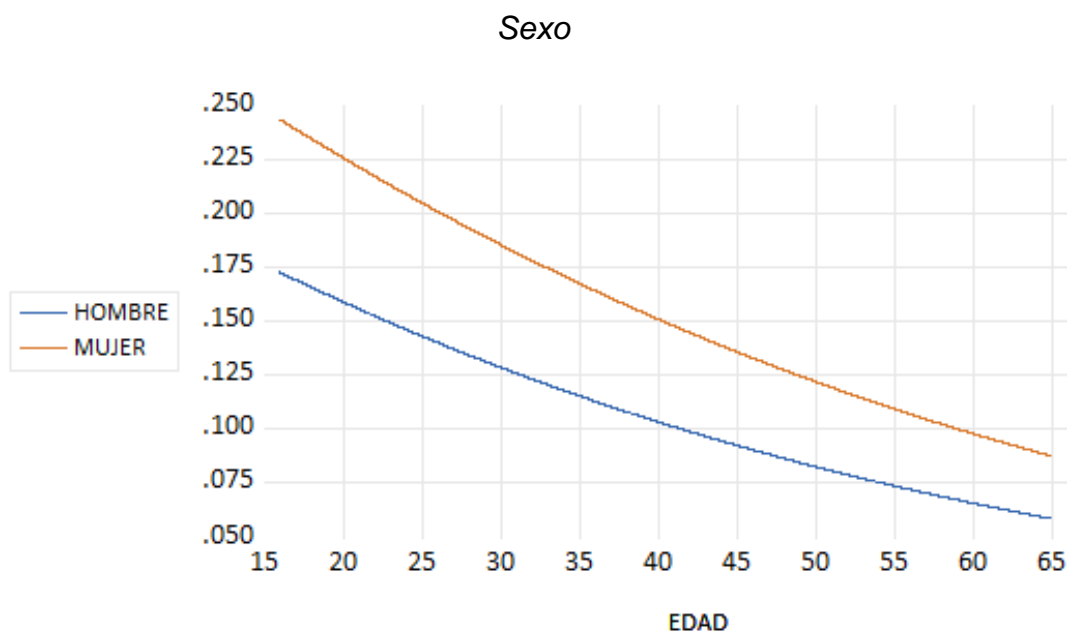


Gráfico 4.6.1: Curva de respuesta en función del sexo.

Fuente: Elaboración propia a partir de la estimación del modelo logit desarrollado en EViews.

En el gráfico 4.6.1 se puede observar cómo existe diferencia entre la probabilidad de encontrarse en situación de desempleo entre hombres y mujeres, siendo este último grupo el que cuenta con mayores posibilidades.

Además, también se observa una tendencia a la reducción de esta desigualdad a medida que aumenta la edad del individuo, es decir, que en los grupos de individuos de mayor edad esta diferencia, aun existente, es significativamente menor.

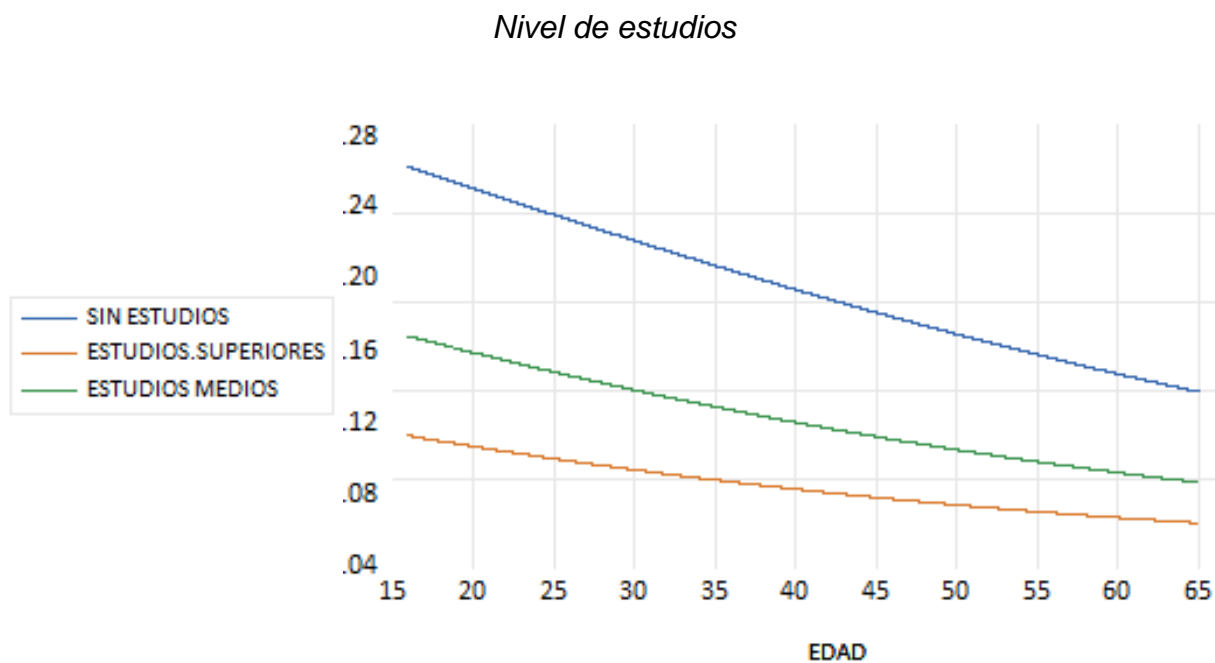


Gráfico 4.6.2: Curva de respuesta en función del nivel de estudios.

Fuente: Elaboración propia a partir de la estimación del modelo logit desarrollado en EViews.

A través del gráfico 4.6.2 se puede ver la clara diferencia existente tanto entre los individuos que poseen la característica estudios superiores frente a los que tienen estudios medios, como entre los que no cuentan con ningún estudio respecto a los anteriores.

Además, también se puede observar la gran distancia probabilística existente entre los individuos analfabetos y sin estudios, frente a aquellos que cuentan con algún tipo de estudio. Como se puede ver reflejado en el gráfico, aquellos individuos que tienen estudios cuentan con más de 15 puntos porcentuales más sobre la probabilidad de tener empleo, frente a aquellos que no tienen estudios.

Nacionalidad

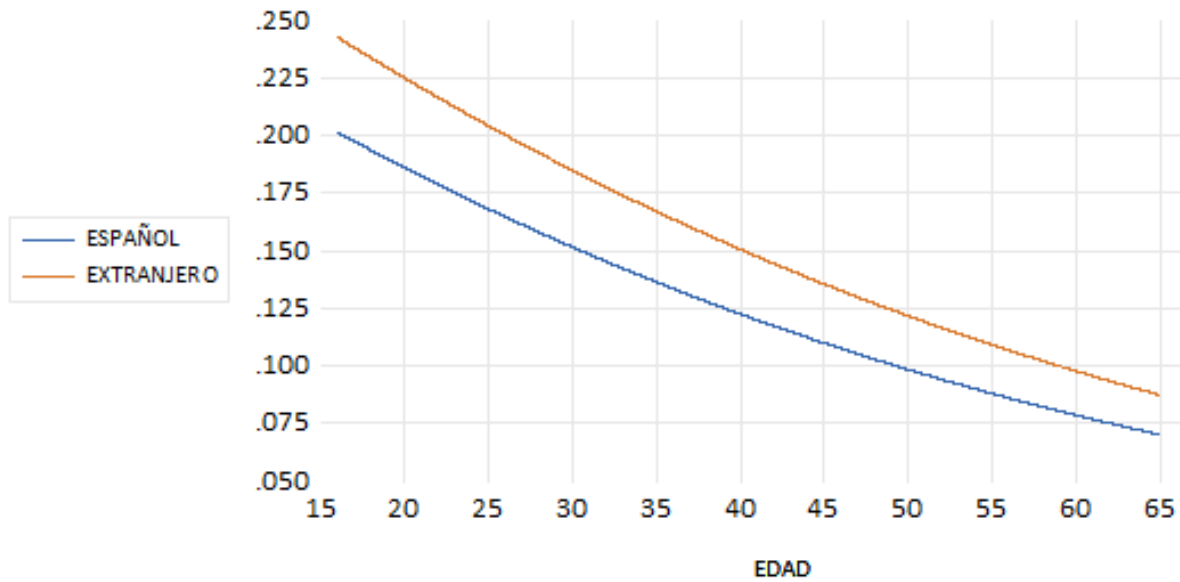


Gráfico 4.6.3: Curva de respuesta en función de la nacionalidad.

Fuente: Elaboración propia a partir de la estimación del modelo logit desarrollado en EViews.

En el gráfico 4.6.3 se puede observar cómo efectivamente existe una diferencia en la probabilidad de estar desempleado, en detrimento de la población extranjera.

Sin embargo, dicha diferencia no es significativamente tan relevante a la hora de explicar la influencia sobre la probabilidad de desempleo como sí lo es el nivel de estudios, por ejemplo.

Estado civil

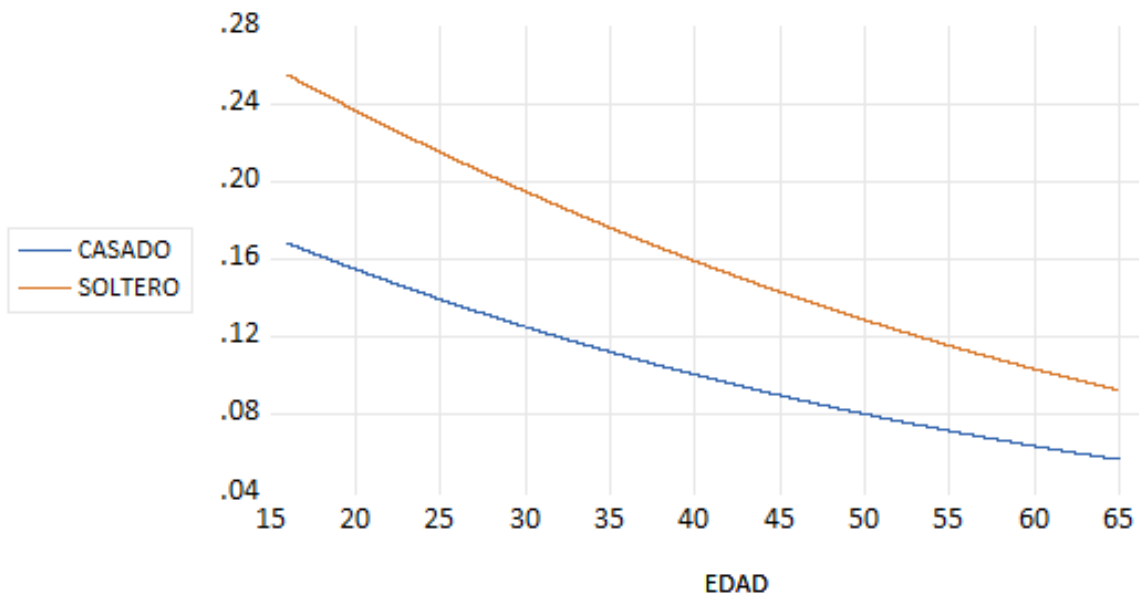


Gráfico 4.6.4: Curva de respuesta en función del estado civil.

Fuente: Elaboración propia a partir de la estimación del modelo logit desarrollado en EViews.

El gráfico 4.6.4 muestra cómo los individuos casados cuentan con mayores probabilidades de tener empleo, frente aquellos solteros.

Al igual que en el resto de gráficos ya analizados, la tendencia hacia un descenso tanto de la diferencia entre poseer o no la característica, como la propia probabilidad del individuo de encontrarse en situación de desempleo, se mantiene, siendo considerablemente superior a mayor edad del individuo.

5. CONCLUSIONES

Ante el objetivo de determinar cómo ciertas características personales influyen en la probabilidad de estar desempleado, se ha llevado a cabo un modelo logit. A la hora de escoger las variables que influirán en la probabilidad de desempleo se ha realizado un estudio tanto de la evolución del estado de paro como de los principales determinantes explicativos del mismo, a saber: edad, sexo, nivel de estudios, nacionalidad, y estado civil.

Tras haber definido el modelo de regresión a emplear, se ha obtenido la influencia resultante de cada variable sobre la probabilidad de desempleo del individuo, siendo todas ellas significativas. De ellas, las que mayores variaciones provocan en el desempleo español son el nivel de estudios, el sexo, y el estado civil, con un 16%, 8,35%, y 10,179% respectivamente, siendo la variable nacionalidad la que resulta ser menos significativa a la hora de influir sobre el mismo.

Sin embargo, y teniendo en cuenta que la característica edad se trata de una variable continua, se han llevado a cabo diferentes escenarios en los que se pudiera observar la influencia de esta variable y del resto. Para ello se han confrontado individuos que reúnen las mismas características a excepción de una de ellas, para así poder contabilizar esa diferencia entre poseer dicha variable o no.

Así se ha llegado a que la mayor probabilidad de encontrarse en desempleo llega al 55%, perteneciente a una mujer de 20 años, soltera, sin estudios, y extranjera. Mientras que por el contrario, un hombre de 60 años, casado, con estudios superiores, y español, cuenta con tan solo con el 3'55% de probabilidades de no tener empleo.

De este modo se ha posicionado la variable edad como la más determinante en cuanto a la variación que provoca sobre la probabilidad de desempleo del individuo, siendo mucho más probable que un individuo joven se encuentre en situación de desempleo frente a un individuo con las mismas características, pero de mayor edad.

Esta tendencia ha resultado mantenerse a lo largo de todas las variables del estudio, lo cual, a través de la realización de las curvas de respuesta, se puede ver esta disminución de la probabilidad de desempleo gráficamente de forma muy notoria.

En cuanto al resto de las variables del estudio, resulta además interesante comentar los resultados obtenidos, no solo en cuanto al porcentaje que aportan a la variación que experimenta el desempleo total, sino también el significado de las diferencias entre que el individuo posea o no la característica.

Empezando por la variable NIVEL DE ESTUDIOS, se ha observado gráficamente cómo la mayor diferencia reside entre los individuos que no poseen ningún estudio frente a los que sí cuentan con algún tipo de estudio, sea del tipo que sea. De este modo, estos últimos individuos cuentan con un porcentaje considerablemente más elevado de probabilidades de tener empleo, frente a los que no. Además, también cabe destacar, que, cómo se ha comentado anteriormente, esta diferencia llega a su máximo a la mínima edad laboral (16 años) del individuo, alcanzando su mínimo a la edad laboral más alta (65 años),

Además, existe una diferencia del 20,46% entre no tener estudios y tener estudios medios, colocando a la variable estudios cómo la segunda variable más significativa del estudio a la hora de explicar el desempleo español, por detrás de la edad.

Pasando a la variable ESTADO CIVIL, se ha observado que los individuos casados tienen menos probabilidad de estar en paro, frente a los individuos solteros, los cuales cuentan con una probabilidad mayor de encontrarse sin trabajo, concretamente un 10,179%.

En cuanto a la variable SEXO, esta muestra la existencia de una brecha laboral estructural de género, en perjuicio de las mujeres, las cuales en posición de

tener las mismas características que un hombre, cuentan con más probabilidades de encontrarse en situación de desempleo (8,35%).

Y por último, la variable NACIONALIDAD, la cual se posiciona en el último puesto en cuanto a su influencia sobre el desempleo se refiere, nos permite observar la diferencia entre individuos españoles y extranjeros, siendo estos últimos los que tienen más probabilidades de estar en paro, aun contando con las mismas características que un nacional. Esto se ve al comparar 2 individuos de 20 años, hombres, solteros, y con estudios superiores, al comprobar que el individuo nacional cuenta con un 3,19% menos de probabilidad de estar encontrarse en situación de desempleo.

6. BIBLIOGRAFÍA Y WEBGRAFÍA

Antonio M. Cardona Álvarez, María Matilde Ceballos Martín, Raúl Pérez Guerra (2019). *Reflexiones sobre el turismo y el empleo en España: el paro, las políticas de empleo y el papel del turismo en los procesos de creación*. Dialnet.

Carrascal, u, y. Gonzalez y b. Rodríguez.: *Análisis Econométrico con E-views*. Ra-Ma. 2001.

EViews 10 Users Guide (2019) [Consulta: 24 de marzo de 2020].

Expansión (2019): “Tasa de desempleo: 2019 Países”. Disponible en: <<https://datosmacro.expansion.com/paro>>. [Consulta: 10 de marzo de 2019].

Gujarati, D.N. (2003) *Econometría* (4 ed.). McGraw-Hill.

Harvey, Andrew C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press.

Instituto Nacional de Estadística (2019): “Encuesta de población activa”. Disponible en: <https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176918&menu=ultiDatos&idp=1254735976595>. [Consulta: 11 de marzo de 2019].

Wooldridge, J.M. (2006) : *Introducción a la Econometría: Un Enfoque Moderno*. Thomson