



TRABAJO DE FIN DE GRADO

CURSO ACADÉMICO 2022/23

**Estimaciones de las Tasas de Paro  
por Provincia y Género en España:  
Análisis de las Disparidades  
Regionales**

UNIVERSIDAD MIGUEL HERNÁNDEZ

FACULTAD DE CIENCIAS SOCIALES Y JURÍDICAS DE ELCHE

GRADO EN ESTADÍSTICA EMPRESARIAL

**Autor:** *Jorge Onofre González*

**Tutor:** *Domingo Morales González*

# Índice

<b>1. Resumen · Abstract</b>	<b>4</b>
<b>2. Introducción</b>	<b>6</b>
<b>3. Metodología Encuesta de Población Activa (EPA)</b>	<b>8</b>
3.1. Introducción . . . . .	8
3.2. Objetivos . . . . .	9
3.3. Necesidad . . . . .	10
3.4. Ámbito geográfico y poblacional . . . . .	11
3.5. Diseño de la encuesta . . . . .	11
3.6. Recogida de la información . . . . .	12
3.7. Definiciones . . . . .	13
3.7.1. Población ocupada . . . . .	13
3.7.2. Población parada o desempleada . . . . .	15
3.7.3. Población inactiva . . . . .	16
<b>4. Descripción del fichero</b>	<b>18</b>
<b>5. Regresión Multinomial</b>	<b>25</b>
5.1. Formulación del modelo . . . . .	25
5.2. Ajuste y diagnóstico del modelo . . . . .	26
5.3. Función multinom de R . . . . .	31
<b>6. Estimadores aplicados a la EPA</b>	<b>32</b>
6.1. Proporción de Parados . . . . .	34
6.2. Proporción de Ocupados . . . . .	35

---

6.3. Tasa de Paro . . . . .	36
<b>7. Resultados estimadores aplicados a la EPA</b>	<b>37</b>
7.1. Estimaciones directas . . . . .	37
7.1.1. Tasas de paro masculinas . . . . .	37
7.1.2. Tasas de paro femeninas . . . . .	39
7.2. Estimaciones modelo multinomial . . . . .	41
7.2.1. Tasas de paro masculinas . . . . .	41
7.2.2. Tasas de paro femeninas . . . . .	42
<b>8. Conclusiones</b>	<b>45</b>
<b>9. Apéndice A - Tabla de coeficientes</b>	<b>48</b>
<b>10. Apéndice B - Tabla de resultados</b>	<b>50</b>
<b>11. Apéndice C - Código R</b>	<b>52</b>
<b>12. Bibliografía</b>	<b>56</b>

## 1. Resumen · Abstract

### *Resumen*

---

*Este estudio tiene como objetivo analizar las tasas de desempleo en España clasificadas por provincia y género. Al examinar detalladamente los datos, se observa que una marcada disparidad entre las provincias del sur y las del norte, mostrando tasas de desempleo significativamente más altas en las primeras. Además, se constata que las mujeres presentan tasas de desempleo más elevadas en comparación con los hombres, lo cual subraya la urgencia de abordar las desigualdades de género en el mercado laboral.*

*Estas tasas son considerablemente superiores a las de otros países de la Unión Europea. Estos hallazgos evidencian la necesidad imperante de implementar políticas efectivas para fomentar la creación de empleo y mejorar las condiciones laborales, con el objetivo de impulsar la economía nacional.*

*En conclusión, los resultados de este estudio enfatizan la importancia de abordar el problema del desempleo en España desde múltiples perspectivas, incluyendo la geográfica y la de género.*

**Palabras clave:** *Tasa de paro - Estimaciones - EPA - Desigualdad laboral.*

### *Abstract*

---

*This study aims to analyze the unemployment rates in Spain classified by province and gender. Upon examining the data in detail, a marked disparity between southern and northern provinces is observed, with significantly higher unemployment rates in the former. Furthermore, it is found that women have higher unemployment rates compared to men, underscoring the urgency of addressing gender inequalities in the labor market.*

*These rates are considerably higher than those in other countries of the European Union. These findings highlight the pressing need to implement effective policies to promote job creation and improve working conditions, with the goal of boosting the national economy.*

*In conclusion, the results of this study emphasize the importance of addressing the issue of unemployment in Spain from multiple perspectives, including geographical and gender considerations.*

**Key words:** *Unemployment rate - Estimations - EPA - Employment inequality.*



## 2. Introducción

En este estudio, nos enfocaremos en explorar las características del mercado laboral en España, centrándonos en la población activa. Nuestro objetivo principal es calcular las tasas de paro, así como las proporciones de personas ocupadas y paradas, desglosadas por provincia y sexo. Para lograrlo, utilizaremos como fuente de referencia los datos de la Encuesta de Población Activa (EPA) del Instituto Nacional de Estadística (INE), correspondientes al primer trimestre del año 2022.

Presentaremos los resultados de manera visual y comprensible, mediante la utilización de mapas y gráficos que representen los hallazgos obtenidos. Mediante el análisis de las estimaciones obtenidas, podremos cuantificar y comprender de manera directa la realidad actual del mercado laboral en España.

Organizaremos la información en diferentes capítulos. En primer lugar, abordaremos los aspectos técnicos de la EPA, ya que es fundamental comprender su metodología y características técnicas, ya que utilizaremos sus datos como base para nuestras estimaciones. Exploraremos cómo se lleva a cabo la encuesta y qué aspectos la definen.

A continuación, nos sumergiremos en el estudio detallado de los datos, centrándonos especialmente en aquellas variables que consideramos más relevantes para nuestra investigación. En esta sección, realizaremos un análisis descriptivo conciso que nos brindará una visión panorámica de los datos en cuestión.

Posteriormente, presentaremos el modelo de regresión multinomial, su la formulación matemática correspondiente y su ajuste a los datos. A continuación, procederemos a realizar la estimación de los indicadores propuestos para el mercado laboral, tanto mediante estimaciones directas como mediante estimaciones basadas en el modelo multinomial. Antes de ello, presentaremos una breve formulación matemática de los estimadores utilizados.

Para llevar a cabo todos los cálculos necesarios en este estudio, emplearemos el lenguaje de programación R en el entorno de RStudio. Finalmente, presentaremos los resultados obtenidos y derivaremos las conclusiones pertinentes a partir de ellos. Además, proporcionaremos una tabla que contiene todos los resultados de las estimaciones de las

tasas de paro, lo que permitirá al lector consultar los datos desglosados por sexo y provincia de manera detallada.

Utilizaremos las variables consideradas más relevantes de un total de 91 para realizar nuestras estimaciones. Con estos datos, emplearemos estimadores de dominio directos para obtener estimaciones sobre la proporción de parados, la proporción de ocupados y las tasas de paro. Todas estas estimaciones se desglosarán por sexo y provincia, lo que nos permitirá cuantificar la situación del mercado laboral español tanto por género como por ubicación geográfica.



## 3. Metodología Encuesta de Población Activa (EPA)

### 3.1. Introducción

La Encuesta de Población Activa (EPA) es una encuesta por muestreo trimestral dirigido a la población que reside en viviendas familiares en el territorio nacional. Su objetivo es obtener información sobre las características de dicha población en relación con el mercado laboral.

La EPA se inició en 1964, con resultados trimestrales hasta finales de 1968. Entre 1969 y 1974, la frecuencia cambió a semestral y luego regresó a trimestral a partir de 1975. Es una encuesta continua en la que las entrevistas se distribuyen de manera uniforme a lo largo de las 52 semanas del año, excluyendo el mes de agosto hasta 1998.

A lo largo de los años, la EPA ha experimentado numerosas modificaciones en su metodología y contenido para adaptarse a los cambios en el mercado laboral español, incorporar avances en la recopilación y análisis de información, y cumplir con las normas internacionales para este tipo de encuestas. Cuando se produce un cambio que afecta a las principales series de resultados, se proporciona un enlace para acceder a estas series.

Entre los cambios destacados se encuentra el realizado en 1987, cuando se modificó el cuestionario de la EPA para cumplir con las recomendaciones de la Conferencia Internacional de Estadísticos del Trabajo de 1982. Esto incluyó la incorporación de variables de la Encuesta de Fuerza de Trabajo de la Unión Europea debido a la entrada de España en las Comunidades Europeas en 1986, y la inclusión de información sobre novedades en el mercado laboral español, como el aumento del trabajo a tiempo parcial y la contratación temporal.

En 2002, se aplicó una nueva definición operativa de desempleo en la EPA, y en 2005 se introdujo un cuestionario más amplio y comprensible, junto con el control centralizado de la recopilación de información a través de encuestas telefónicas asistidas por ordenador. Además, a partir de ese año, algunas variables de la encuesta se recopilan exclusivamente en una submuestra anual representativa de la situación promedio del año, en lugar de trimestralmente. Estas variables se conocen como “variables de submuestra anual”.

En 2021, la encuesta se ha adaptado a la nueva normativa europea que regula las encuestas dirigidas a la población, específicamente la Encuesta de Fuerza de Trabajo. Esto incluye varios reglamentos del Parlamento Europeo y de la Comisión. Muchos de los requisitos de la nueva normativa ya se cumplían en la EPA antes de 2021. Las cifras actuales de la encuesta siguen la metodología establecida en 2005 con adaptaciones basadas en este cambio normativo y en resoluciones de la OIT de 2013 y 2018.

Además de los cambios mencionados, también se prevén modificaciones en el diseño y la ejecución normal de la encuesta, como la actualización periódica de la muestra y su rotación.

### 3.2. Objetivos

La Encuesta de Población Activa (EPA) tiene como objetivo principal proporcionar información sobre la actividad económica en términos de su componente humano. Su enfoque está en ofrecer datos sobre las categorías poblacionales clave relacionadas con el mercado laboral, como ocupados, parados, activos e inactivos, y clasificar estas categorías según diversas características. Además, permite crear series temporales consistentes de resultados. Importante destacar que las definiciones y criterios utilizados son coherentes con los estándares establecidos por organismos internacionales en temas laborales, lo que posibilita la comparación con datos de otros países.

La EPA proporciona resultados detallados a nivel nacional. Para las comunidades autónomas y provincias, se ofrece información sobre las principales características con el nivel de desglose permitido por el coeficiente de variación de los estimadores.

### 3.3. Necesidad

Existen diversas fuentes estadísticas que proporcionan información sobre la fuerza de trabajo, pero cada una de ellas tiene limitaciones que hacen necesaria una encuesta específica, como la Encuesta de Población Activa (EPA).

Los Censos de Población ofrecen información sobre la fuerza de trabajo, pero presentan inconvenientes como el tiempo transcurrido desde su realización, la autocompletación de los cuestionarios por parte de los entrevistados, altos costos y la demora en la obtención de resultados debido al gran volumen de la operación.

Las Encuestas de Salarios y las Encuestas Industriales solo brindan información sobre los asalariados y no cubren todas las ramas de actividad.

El Paro Registrado y la afiliación a la Seguridad Social, al ser estadísticas basadas en registros administrativos, están sujetos a normas legales variables, ofrecen información limitada sobre una parte específica de la población estudiada y no permiten obtener series homogéneas.

Por otro lado, una encuesta de población activa presenta diversas ventajas:

- Se puede realizar de forma continua según la periodicidad deseada.
- Permite un enfoque directo y detallado en relación con la fuerza laboral.
- Cuenta con entrevistadores especializados para la recopilación de datos.
- Los resultados se obtienen rápidamente debido al muestreo.
- Se emplean definiciones y tratamientos de información uniformes en las sucesivas realizaciones de la encuesta, lo que genera series homogéneas de resultados y proporciona información a nivel nacional y territorial.

El principal inconveniente de la EPA es su naturaleza de encuesta por muestreo, lo que limita la capacidad de brindar información detallada con la máxima desagregación posible. A medida que se desglosa la información a niveles más específicos, como el número de activos por divisiones de actividad económica en cada provincia, la fiabilidad disminuye debido al error de muestreo.

### 3.4. Ámbito geográfico y poblacional

La Encuesta de Población Activa (EPA) abarca todo el territorio nacional, excepto Ceuta y Melilla, cuya incorporación se produjo en 1988.

La EPA se dirige a la población que reside en viviendas familiares principales, es decir, aquellas utilizadas durante todo el año o la mayor parte de él como residencia habitual o permanente. No se incluyen hogares colectivos ni viviendas secundarias o de temporada.

En el caso de familias que utilizan dos viviendas de manera habitual, se considera como vivienda principal aquella que la familia reconozca como tal, y se encuesta a todos los miembros en esa vivienda. Además, la denominación de vivienda familiar también incluye a aquellas en las que los componentes no tienen vínculos familiares.

A partir de 2021, se utiliza el criterio de residencia y el compartir un presupuesto común como definición de hogar en todas las encuestas de la población. Si en una vivienda familiar hay más de un hogar, se encuestan todos ellos.

### 3.5. Diseño de la encuesta

La Encuesta de Población Activa (EPA) utiliza un método de **muestreo bietápico con estratificación** en las unidades de primera etapa para cada provincia.

1. En la primera etapa, son las secciones censales. Estas secciones son áreas geográficas delimitadas según la Ley Electoral, con una población de entre 500 y 2000 electores.
2. En la segunda etapa, se seleccionan las viviendas familiares principales sin realizar submuestreo, lo que implica entrevistar a todas las personas que residen en esas viviendas.

Una vez que una vivienda ha sido seleccionada para formar parte de la encuesta, se mantiene en ella durante seis trimestres consecutivos, donde al final será sustituida por otra de la misma sección.

### 3.6. Recogida de la información

La captura de información para la EPA se lleva a cabo de la siguiente manera:

- Las primeras entrevistas se realizan mediante visitas personales asistidas por ordenador (método CAPI). Los entrevistadores de las 52 delegaciones provinciales del INE realizan estas entrevistas semanalmente.
- Las segundas y sucesivas entrevistas (de la 2ª a la 6ª) se realizan mediante llamadas telefónicas asistidas por ordenador (método CATI). Estas entrevistas se llevan a cabo desde siete centros CATI ubicados en las delegaciones provinciales de Barcelona, Cádiz, A Coruña, Sevilla, Valencia, Vizcaya y Servicios Centrales.
- Desde el año 2020, también se permite que la encuesta se complete a través de Internet (CAWI).

En las delegaciones provinciales, además, se recopila información de aquellas viviendas que, a pesar de estar en las entrevistas sucesivas, prefieren continuar colaborando mediante visitas personales. Después de recoger la información, se transmite a los Servicios Centrales para su posterior procesamiento.

(\*) NOTA:

<b>CAPI</b>	Computer Assisted Personal Interviewing
<b>CATI</b>	Computer Assisted Telephone Interviewing
<b>CAWI</b>	Computer Assisted Web Interviewing

### 3.7. Definiciones

Las siguientes definiciones se basan en las recomendaciones de la Organización Internacional del Trabajo (OIT) en conferencias de Estadísticos del Trabajo. Estas definiciones se refieren al concepto nacional y no al interior, según el Sistema Europeo de Cuentas Nacionales y Regionales (SEC-95). Esto se debe a la imposibilidad de recopilar información sobre la población que trabaja en España pero reside en el extranjero, ya que la encuesta se dirige a la población que vive en viviendas familiares dentro del territorio nacional.

#### 3.7.1. Población ocupada

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

##### A) Personas con empleo por cuenta ajena o asalariadas:

- Se consideran personas con trabajo a aquellas que han trabajado al menos una hora a cambio de un salario o retribución en metálico o especie durante la semana de referencia, incluso de forma esporádica u ocasional.
- También se incluyen las personas que tienen empleo pero no trabajaron durante la semana de referencia debido a enfermedad, accidente, vacaciones, licencia de maternidad/paternidad u otros motivos, siempre y cuando mantengan un vínculo estrecho con su empleo.
- Las personas con contratos fijos discontinuos, trabajadores estacionales o en espera de incorporación a un nuevo empleo se consideran con empleo en los períodos en los que trabajan. En el resto del tiempo, se consideran ocupados solo si realizan tareas relacionadas con su empleo de manera regular.
- En casos de excedencia por maternidad o paternidad pagada, también se consideran ocupados.

- Las personas ausentes por otros motivos distintos a los anteriores se consideran ocupadas si tienen previsto reincorporarse a su empleo en los próximos tres meses. Durante la pandemia de COVID-19, las personas ausentes por expediente de regulación temporal de empleo se consideran que mantienen un vínculo fuerte con su empleo si reciben una prestación superior al 50 % del salario.
- Los aprendices y estudiantes que reciben una remuneración se clasifican como personas asalariadas y se consideran trabajando o sin trabajar en función de las mismas condiciones que las demás personas con empleo por cuenta ajena.

#### **B) Personas con actividad por cuenta propia:**

- Se consideran personas con actividad por cuenta propia aquellas que han trabajado al menos una hora a cambio de un beneficio o ganancia familiar, ya sea en metálico o en especie, durante el período de referencia, incluso de forma esporádica u ocasional.
- También se incluyen las personas que tenían que realizar algún trabajo a cambio de un beneficio o ganancia familiar, pero estuvieron temporalmente ausentes debido a razones como enfermedad, accidente, vacaciones, jornada de verano u horarios variables, flexibles o similares. En casos de otras ausencias como fiestas o mal tiempo, la duración debe ser de tres meses o menos.
- Se consideran personas con actividad por cuenta propia a los empresarios, trabajadores independientes, miembros de cooperativas que trabajan en ellas y trabajadores familiares no remunerados.
- No se consideran personas con actividad por cuenta propia a aquellas que se ocupan de su hogar sin remuneración, prestan servicios sociales no remunerados o ejercen actividades fuera del ámbito de las actividades económicas.
- Los trabajadores estacionales por cuenta propia y los trabajadores estacionales u ocasionales familiares no remunerados en la temporada de menor actividad que no hayan trabajado en la semana de referencia ni realicen actividades relacionadas con el negocio estacional no se consideran ocupados por cuenta propia.

### 3.7.2. Población parada o desempleada

Se considerarán como personas en situación de desempleo aquellas que cumplan simultáneamente con las siguientes condiciones, siempre y cuando tengan 16 años o más:

1. No tener empleo remunerado ni como empleado ni como trabajador autónomo durante la semana de referencia.
2. Estar activamente buscando empleo, lo cual implica haber realizado esfuerzos concretos para encontrar un trabajo asalariado o haber emprendido acciones para establecer un negocio propio durante el mes anterior.
3. Estar disponibles para comenzar a trabajar en un plazo máximo de dos semanas a partir del domingo de la semana de referencia.

Además, también se considerarán como personas en situación de desempleo aquellas que cumplan con los siguientes requisitos durante la semana de referencia:

1. Estar sin empleo y disponibles para trabajar.
2. No buscar empleo debido a que ya han conseguido uno al cual se incorporarán en un plazo de hasta tres meses a partir de dicha semana. En este caso, no es necesario cumplir con el criterio de búsqueda activa de empleo.
3. Las personas que están fuera del mercado laboral y no son consideradas ocupadas según lo establecido en el punto 3.7.1, pero que planean reincorporarse a su empleo después de tres meses, se considerarán desempleadas si han buscado trabajo y están disponibles para desempeñarlo.

Si una persona cumple con las condiciones mencionadas anteriormente, se considerará desempleada, aunque pueda combinar esta situación con otras, como ser estudiante o realizar tareas domésticas.

Las personas en situación de desempleo se dividen en dos categorías: aquellos que buscan su primer empleo y aquellos que han trabajado anteriormente.

### 3.7.3. Población inactiva

La categoría de la población económicamente inactiva incluye a todas las personas de 16 años o más que no se clasifican como ocupadas ni desempleadas ni pertenecen a una categoría específica aparte durante la semana de referencia. Esta categoría comprende las siguientes situaciones:

- Personas que se ocupan de su hogar: aquellas que no realizan ninguna actividad económica y se dedican al cuidado de sus propios hogares, como amas de casa y otros familiares encargados del cuidado del hogar y de los niños.
- Estudiantes: personas que no llevan a cabo ninguna actividad económica y reciben educación formal en cualquier nivel.
- Jubilados o prejubilados: personas que han tenido una actividad económica previa y la han dejado debido a su edad u otras circunstancias, recibiendo una pensión o ingresos de prejubilación relacionados con su actividad anterior.
- Personas que reciben una pensión distinta de la de jubilación o prejubilación.
- Personas que realizan trabajos sociales, actividades benéficas, etc., sin recibir remuneración (excepto aquellas que son ayudas familiares).
- Personas incapacitadas para trabajar.
- Otras situaciones: personas que no realizan ninguna actividad económica y reciben ayuda pública o privada, así como todas aquellas que no se encuentran incluidas en ninguna de las categorías anteriores, como los rentistas.

Dentro de los inactivos, se distingue un grupo especial denominado “activos potenciales”, que son aquellos inactivos que están disponibles para trabajar pero no están buscando empleo. Se excluye a aquellos que no buscan empleo debido a la jubilación, la educación, la enfermedad, el cuidado de personas dependientes u otras responsabilidades personales o familiares.

Dentro de los activos potenciales, hay un subgrupo de interés conocido como “desanimado”, que son personas sin empleo, disponibles para trabajar en un período de dos semanas, pero que no han buscado empleo en las últimas cuatro semanas debido a la falta de oportunidades laborales adecuadas.



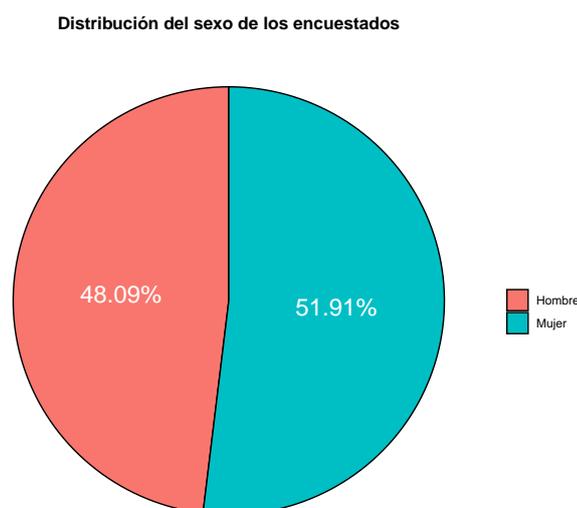
## 4. Descripción del fichero

En esta sección, presentaremos un breve análisis descriptivo de los datos correspondientes al primer trimestre de 2022 según la Encuesta de Población Activa (EPA).

En primer lugar, nos enfocaremos en los aspectos más relevantes relacionados con nuestras variables de interés, las cuales utilizaremos para crear los dominios de estudio. Inicialmente, contamos con un total de 131.808 observaciones correspondientes a 91 variables, aunque solo algunas de ellas serán consideradas de interés para nuestro análisis.

Entre estas variables, destacamos el sexo, el cual puede ser masculino o femenino, así como la variable provincia, que puede tomar valores correspondientes a las 50 provincias del país y las 2 ciudades autónomas. A partir de estas dos variables, crearemos lo que denominamos “áreas”, que son las distintas combinaciones entre las provincias y los géneros. En otras palabras, tendremos un área para cada provincia cruzada con cada sexo. En total, contaremos con 104 áreas definidas en nuestro estudio.

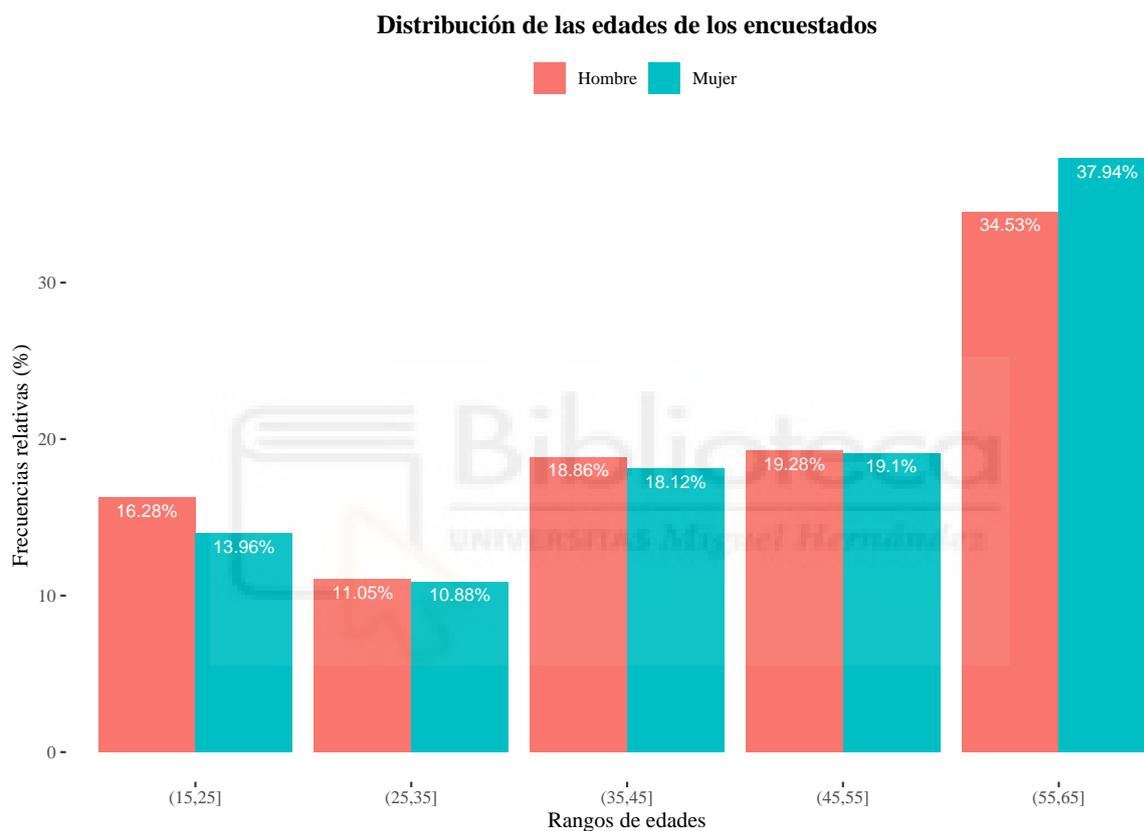
Sin embargo, antes de realizar cualquier estimación, es importante familiarizarnos con estas variables y otras relevantes para el año de referencia. Comenzaremos analizando el sexo de los encuestados en la EPA, visualizando en la **Figura 1** la distribución de éstos mediante un gráfico de tipo circular para tener una visión general de la proporción de hombres y mujeres en la muestra.



**Figura 1:** Distribución de los sexos en los datos de 2022-T1

Durante el periodo de referencia, se puede apreciar que el número de mujeres que participan en la encuesta es ligeramente mayor que el de hombres.

En relación a la edad de los encuestados en la EPA, podemos dividirla en cinco categorías diferentes según los intervalos de edad (15, 25], (25, 35], (35, 45], (45, 55], (55, 65]. La **Figura 2** muestra las frecuencias relativas para hombres y mujeres por intervalos de edad, permitiendo observar la distribución de edades de los encuestados en la muestra.



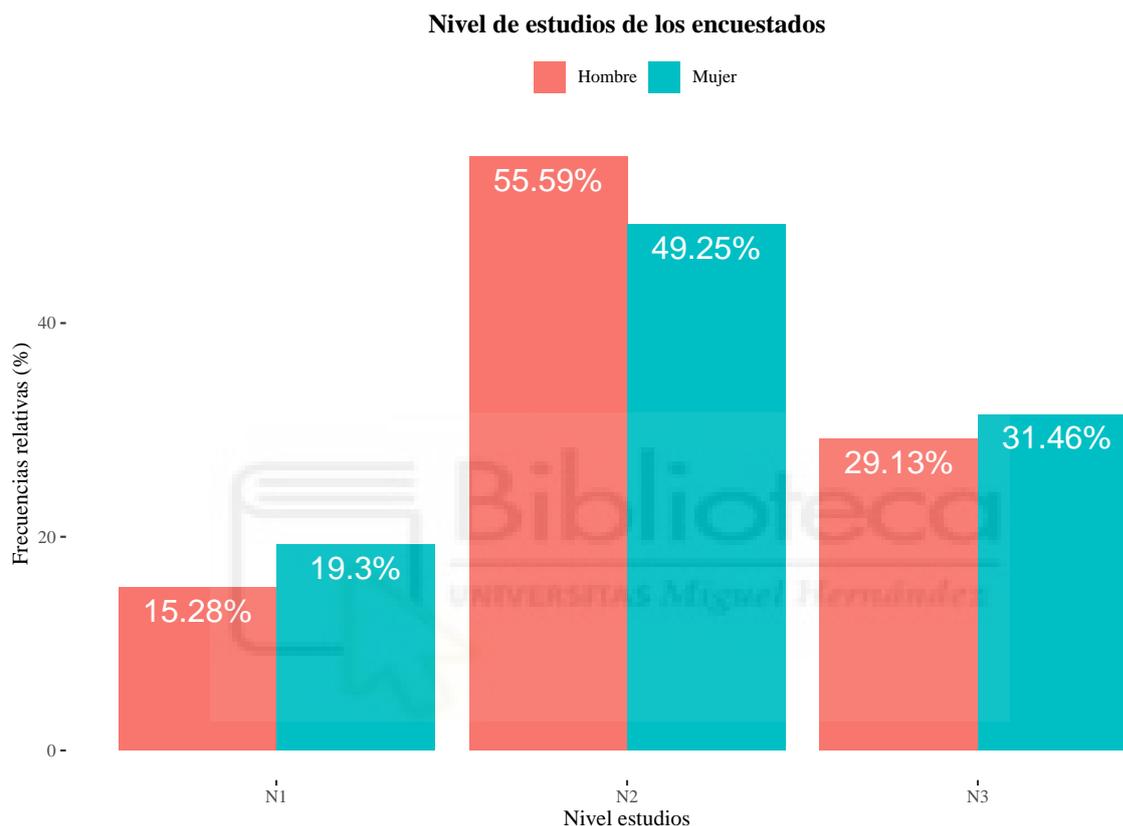
**Figura 2:** Distribución de las edades en los datos de 2022-T1

Podemos observar que a medida que la edad de los encuestados aumenta, la frecuencia relativa también aumenta, lo que indica una mayor representación de las personas mayores en nuestra muestra en comparación con las más jóvenes. Este patrón de aumento gradual se refleja de manera casi lineal en nuestro gráfico de barras.

En cuanto a las diferentes franjas de edad, encontramos que la más numerosa abarca desde los 56 hasta los 65 años, mientras que la franja menos numerosa corresponde al rango de edad entre los 25 hasta los 35 años. Estos resultados nos brindan información

sobre la distribución de edades en nuestra muestra y nos ayudan a comprender mejor la composición de los encuestados según su edad.

En relación al nivel de estudios de los encuestados en la EPA, hemos categorizado los datos en tres niveles diferentes: N1, N2 y N3. En la **Figura 3** se presentan las frecuencias relativas para hombres y mujeres en cada uno de estos niveles.



**Figura 3:** Distribución del nivel de estudio de los encuestados en los datos de 2022-T1

Podemos observar que más de la mitad de los encuestados se encuentran en el nivel de estudios N2, el cual abarca a individuos con educación secundaria (primera y segunda etapa de educación secundaria con orientación general o profesional). Este nivel es el más representado en nuestra muestra.

Alrededor del 30 % de los encuestados tienen un nivel de estudios N3, lo que indica que poseen educación superior. Este grupo está compuesto por individuos con estudios universitarios o equivalentes.

Por otro lado, menos del 20 % de los encuestados se encuentran en el nivel de

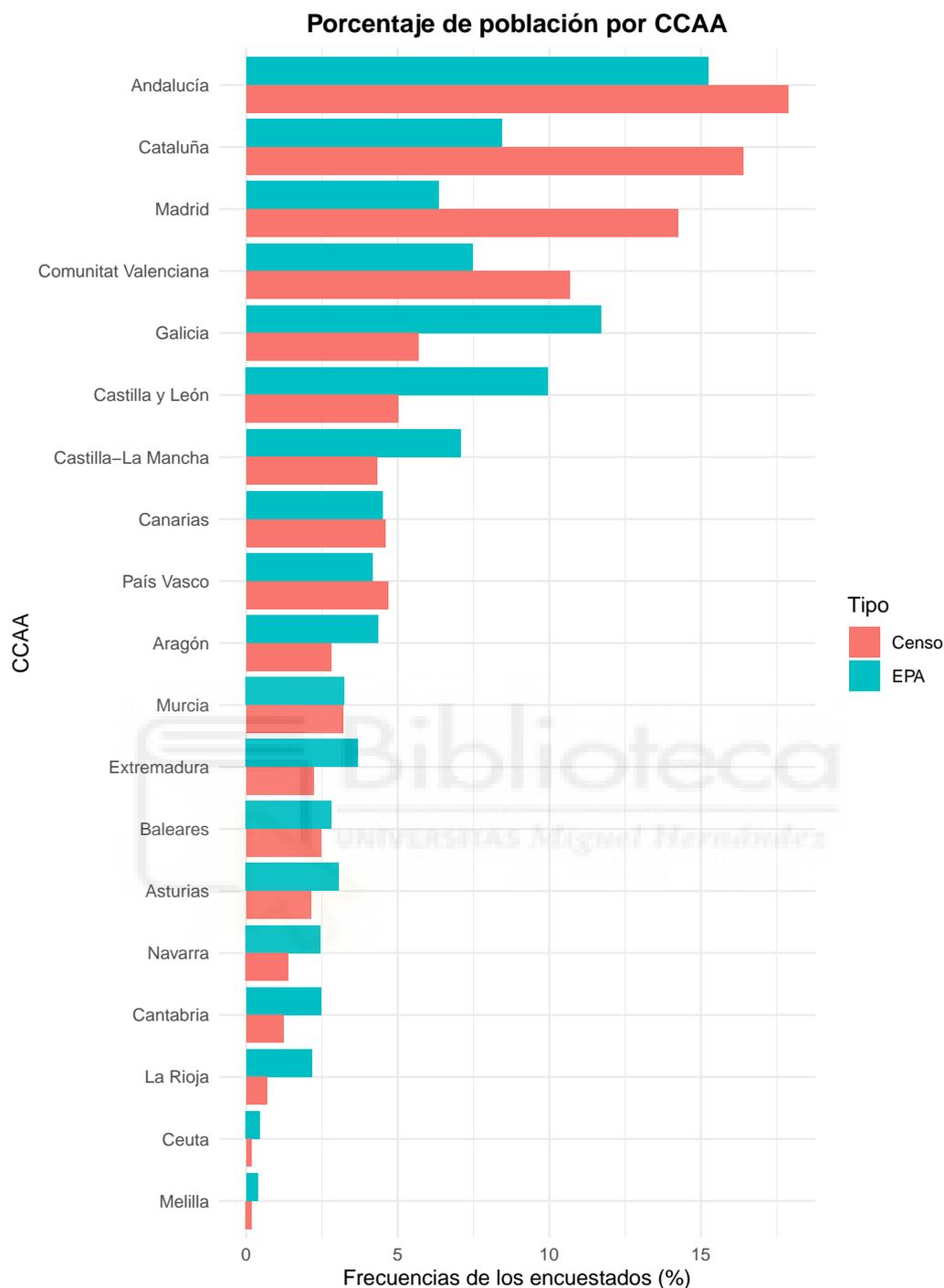
estudios N1, que engloba a aquellos con un nivel educativo básico, como analfabetos, personas con educación primaria incompleta o que solo han completado la educación primaria.

Estos resultados nos brindan una visión general de la distribución del nivel de estudios en nuestra muestra y nos permiten comprender la composición educativa de los encuestados en la EPA.

En resumen, los resultados muestran disparidades en la distribución de los niveles de estudios según el sexo. Las mujeres están más representadas en el nivel educativo básico (N1), mientras que los hombres tienen una mayor presencia en el nivel de educación secundaria (N2). Ambos sexos están considerablemente presentes en el nivel de educación superior (N3). Estas conclusiones ofrecen una perspectiva relevante para comprender las características educativas de la población en función del sexo.

La metodología de muestreo utilizada en la EPA puede tener un impacto en la representación de las distintas comunidades autónomas. Si el diseño de muestreo no se realiza de manera adecuada o no se ajusta correctamente a la distribución demográfica de las regiones, es posible que algunas áreas estén subrepresentadas mientras que otras estén sobrerrepresentadas en la muestra.

Es importante analizar la distribución de los encuestados según la comunidad autónoma a la que pertenecen. La **Figura 4** muestra las frecuencias relativas de los encuestados por comunidad autónoma, comparando su representación en la EPA con la representación de cada comunidad autónoma en el censo proporcionado por el INE. Esto nos permite evaluar si existen discrepancias entre la muestra y la población total en términos de representatividad por comunidad autónoma.



**Figura 4:** Comparación de la distribución de los encuestados y el censo real según la CCAA.

Se puede observar que en la Encuesta de Población Activa (EPA) las comunidades autónomas más representadas son Andalucía, Galicia, Castilla y León, Cataluña y Comunitat Valenciana. Por otro lado, las comunidades autónomas menos representadas en la

encuesta son Navarra, Cantabria, La Rioja, Ceuta y Melilla.

Al comparar las cifras de las comunidades autónomas según su participación en la EPA y su presencia en el censo proporcionado por el INE, se aprecia que comunidades como Andalucía, Cataluña, Madrid o Comunitat Valenciana están subrepresentadas en la muestra. Por el contrario, comunidades como La Rioja, Cantabria, Navarra, Asturias, Extremadura, Castilla y León o Galicia están sobrerrepresentadas en la muestra.

En resumen, si la metodología de muestreo utilizada en la EPA no se ajusta adecuadamente a la distribución demográfica de las regiones, es probable que se generen diferencias en la representación de las comunidades autónomas en la muestra. Esto podría dar lugar a resultados sesgados y afectar la interpretación de los datos.

Otra variable de gran relevancia para nuestro estudio, que además servirá como variable de respuesta, es la clasificación de los encuestados según su relación con la actividad económica, siguiendo los criterios establecidos por la Organización Internacional del Trabajo (OIT). Para facilitar el análisis, hemos realizado una recodificación de la variable original (AOI), que constaba de 7 opciones laborales y la hemos reducido a 3 opciones: ocupados, parados e inactivos. Estas categorías nos permitirán calcular las estimaciones propuestas anteriormente.

A continuación, detallaremos la recodificación que hemos llevado a cabo, identificando los valores que hemos eliminado y los nuevos niveles que hemos creado para nuestra variable:

- Ocupados:
  - Ocupados subempleados por insuficiencia de horas.
  - Resto de ocupados.
- Parados:
  - Parados que buscan primer empleo.
  - Parados que han trabajado antes.

- Inactivos:
  - Inactivos 1 (desanimados).
  - Inactivos 2 (junto con los desanimados forman los activos potenciales).
  - Inactivos 3 (resto de inactivos).

En la **Tabla 1** se muestran las frecuencias relativas de la categoría laboral de los encuestados, distinguiendo entre aquellos que están ocupados, parados o inactivos, desglosados por sexo:

	OCUPADOS	PARADOS	INACTIVOS
HOMBRES	51,92 %	6,59 %	41,49 %
MUJERES	41,94 %	7,35 %	50,73 %

**Tabla 1:** Frecuencias relativas de ocupados, parados e inactivos por sexo.

En la tabla anterior, se puede observar una diferencia notable en el porcentaje de personas ocupadas entre hombres y mujeres, siendo este más alto en hombres que en mujeres. Además, se destaca que el porcentaje de personas inactivas es considerablemente más bajo en hombres que en mujeres.

Adicionalmente, se aprecia que en el caso de las mujeres, hay un mayor porcentaje de personas desempleadas en comparación con los hombres. Esto indica que, en la muestra analizada, las mujeres presentan una proporción más alta de desempleo en relación a los hombres.

Estos resultados revelan disparidades en la situación laboral entre hombres y mujeres, lo cual puede ser objeto de análisis en términos de igualdad de género y oportunidades laborales.

Por último, cabe destacar la importancia del factor de elevación en el cálculo de las estimaciones. Este factor representa el número de personas en la población que cada individuo de la muestra representa. En otras palabras, es una medida de la representatividad de cada encuestado en relación a la población total.

## 5. Regresión Multinomial

La **regresión multinomial** es una generalización del modelo de regresión logística que se utiliza cuando la variable dependiente,  $Y$ , tiene tres o más categorías. Se asume que  $Y$  sigue una distribución multinomial, lo que implica que cada categoría tiene su propia probabilidad asociada. Este enfoque estadístico nos permite analizar las relaciones entre las variables independientes y las categorías de  $Y$ . Es una herramienta útil para comprender las asociaciones y patrones en datos con múltiples categorías de respuesta.

### 5.1. Formulación del modelo

Sea  $Y$  una variable de respuesta categórica con  $R$  categorías, y sean  $\pi_1, \pi_2, \dots, \pi_R$  las probabilidades asociadas a cada categoría, donde  $\sum_{r=1}^R \pi_r = 1$ . En el modelo de regresión multinomial, se elige una de las categorías como referencia (por ejemplo, la última categoría,  $R$ ) y se establece un modelo logit en relación a esta categoría de referencia.

$$\log \left( \frac{\pi_r}{\pi_R} \right) = \alpha_r + \sum_{k=1}^K \beta_{rk} X_{rk}, \quad r = 1, \dots, R - 1 \quad (1)$$

El modelo consta de  $R - 1$  ecuaciones, cada una con sus propios parámetros. Estos parámetros representan el efecto relativo de cada categoría en comparación con la categoría de referencia. Utilizando la ecuación 1 y tomando dos categorías arbitrarias,  $a$  y  $b$ , se puede expresar que:

$$\begin{aligned} \log \left( \frac{\pi_a}{\pi_b} \right) &= \log \left( \frac{\frac{\pi_a}{\pi_R}}{\frac{\pi_b}{\pi_R}} \right) \\ &= \log \left( \frac{\pi_a}{\pi_R} \right) - \log \left( \frac{\pi_b}{\pi_R} \right) \\ &= (\alpha_a + \beta_{1a}X_1 + \dots + \beta_{ka}X_K) - (\alpha_b + \beta_{1b}X_1 + \dots + \beta_{kb}X_K) \\ &= (\alpha_a - \alpha_b) + (\beta_{1a} - \beta_{1b})X_1 + \dots + (\beta_{ka} - \beta_{kb})X_K \end{aligned} \quad (2)$$

Así se obtiene la ecuación logit para la categoría  $a$  en relación en relación a cualquier otra categoría  $b$ ,  $\alpha = (\alpha_a - \alpha_b)$ ,  $\beta_1 = (\beta_{1a} - \beta_{1b})$ ,  $\dots$  y  $\beta_k = (\beta_{ka} - \beta_{kb})$ .

Los programas estadísticos ajustan simultáneamente todas las ecuaciones y los errores obtenidos para los parámetros estimados son más pequeños que los de la regresión. Los parámetros del modelo se estiman utilizando el método de máxima verosimilitud, similar al caso de la regresión logística. Una vez que se han calculado estos parámetros, se pueden obtener las probabilidades predichas para cada categoría resolviendo la ecuación 1 para  $\pi_r$ :

$$\hat{\pi}_r = \frac{\exp\{\alpha_{ri} + \sum_{i=1}^n \beta_r X_i\}}{1 + \sum_{h=1}^{R-1} \exp\{\alpha_h + \sum_{i=1}^n \beta_{hi} X_i\}}, \quad r = 1, \dots, R \quad (3)$$

## 5.2. Ajuste y diagnóstico del modelo

Procederemos a ajustar el modelo multinomial a nuestros datos. El modelo nos permitirá explorar cómo diferentes factores se relacionan con las diferentes categorías de nuestra variable de interés.

La variable que utilizaremos como variable respuesta, es la clasificación de los encuestados según su relación con la actividad económica (AOI). Como vimos anteriormente en la página 23, realizamos una recodificación de la variable original, reduciéndola a 3 categorías: ocupados, parados e inactivos. Así pues, cada categoría tendrá una probabilidad asociada a cada una:  $\pi_1$ ,  $\pi_2$  y  $\pi_3$ .

Usaremos la función `multinom` de la librería `nnet` de R para ajustare el modelo de regresión multinomial utilizando todas nuestra variables de interés, es decir:

$$Y \sim \text{EDAD1} + \text{PROVn} + \text{SEXOn} + \text{FORMAn}$$

En el Apéndice se encuentra disponible la **Tabla 2**, donde se presentan las estimaciones de los parámetros, los errores estándar y los intervalos de confianza al 95 %. Se observa que todas las variables explicativas son significativas.

Al interpretar los parámetros del modelo ajustado, nos adentraremos en el significado de cada coeficiente para comprender las relaciones entre las variables y su impacto en la variable dependiente. Se puede concluir lo siguiente:

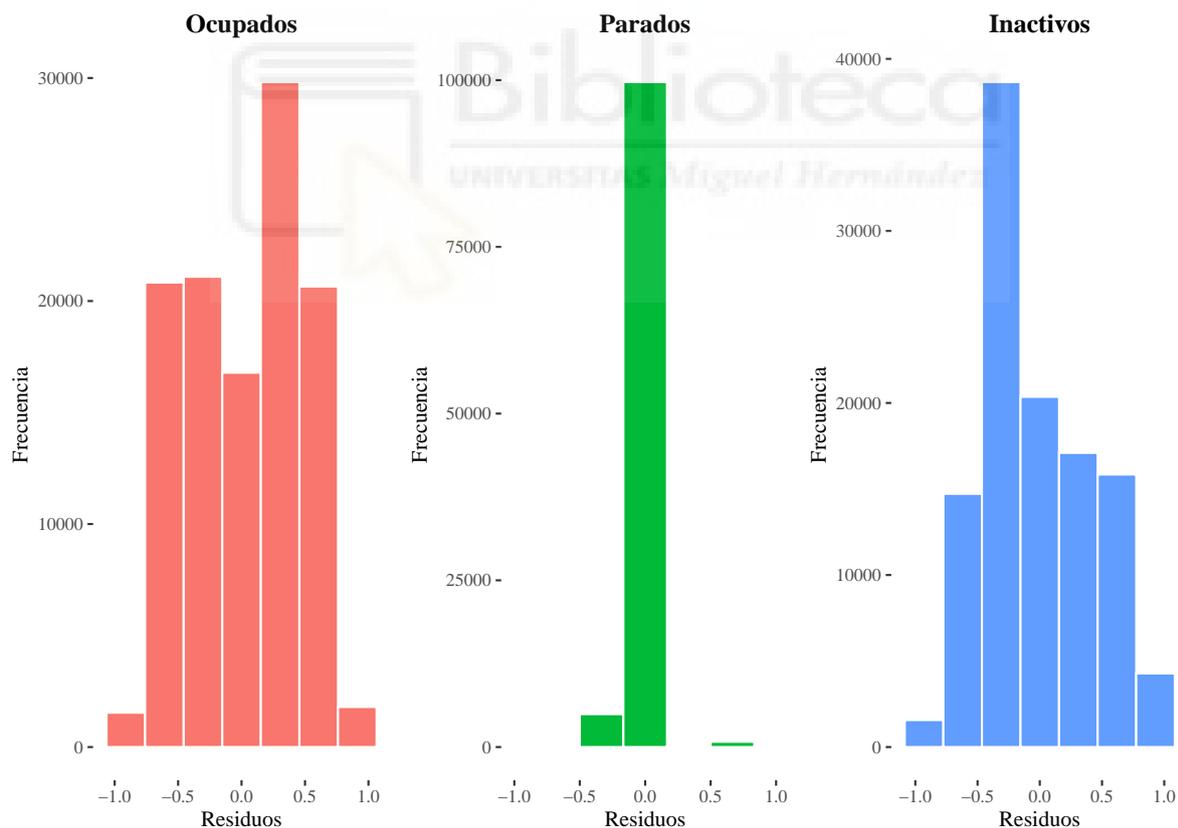
- Las mujeres tienden a tener una tasa de desempleo más alta en comparación con los hombres. Un coeficiente positivo de  $+0,39$  para la variable sexo femenino indica que la pertenencia al género femenino está asociada con un aumento en la tasa de desempleo.
- Aquellos con el nivel de estudios más alto tienen una probabilidad aún menor de estar desempleados en comparación con aquellos con el nivel de estudios más bajo.
  - El coeficiente de  $-0,94$  para la variable N2 indica que, en comparación con el nivel de estudios más bajo (N1), tener un nivel de estudios intermedio (N2) está asociado con una disminución de  $0.94$  en la tasa de desempleo.
  - Por otro lado, el coeficiente de  $-1,74$  para la variable N3 indica que, en comparación con el nivel de estudios más bajo (N1), tener el nivel de estudios más alto (N3) está asociado con una disminución de  $1.74$  en la tasa de desempleo.
- Aquellas provincias con los coeficientes más altos tienden a tener una tasa de desempleo más alta en comparación con aquellas provincias con valores más bajos.
  - Para las provincias de Lérida, Huesca y Gipúzcoa, los coeficientes negativos ( $-0,51$ ,  $-0,45$  y  $-0,32$  respectivamente) indican que estas provincias tienen una disminución en la tasa de desempleo.
  - Por otro lado, para las provincias de Ceuta, Cádiz y Melilla, los coeficientes positivos ( $+1,15$ ,  $+1,10$  y  $+0,90$  respectivamente) indican que estas provincias tienen un aumento en la tasa de desempleo.

A continuación, vamos a estudiar la diagnosis del modelo para evaluar su desempeño y validar su adecuación. La diagnosis del modelo nos brinda información crucial sobre la calidad de ajuste y la fiabilidad de los resultados obtenidos. Exploraremos diversos aspectos de la diagnosis, incluyendo la evaluación de la normalidad de los residuos y la

detección de valores atípicos. Al examinar estos aspectos clave, podremos obtener una visión más completa del modelo y tomar decisiones fundamentadas en cuanto a su validez y confiabilidad.

En primer lugar, en la **Figura 5** se puede estudiar la normalidad de los residuos del modelo ajustado mediante el histograma de frecuencias de los residuos brutos para cada categoría. Esta análisis nos permitirá evaluar si los residuos siguen una distribución normal.

Los datos de residuos han sido divididos en tres categorías: ocupados, parados e inactivos. Observaremos la forma de los histogramas para cada categoría y analizaremos si se asemejan a una distribución normal o presentan desviaciones significativas. Esta evaluación nos proporcionará una comprensión más profunda de la adecuación del modelo ajustado y la validez de los resultados obtenidos.

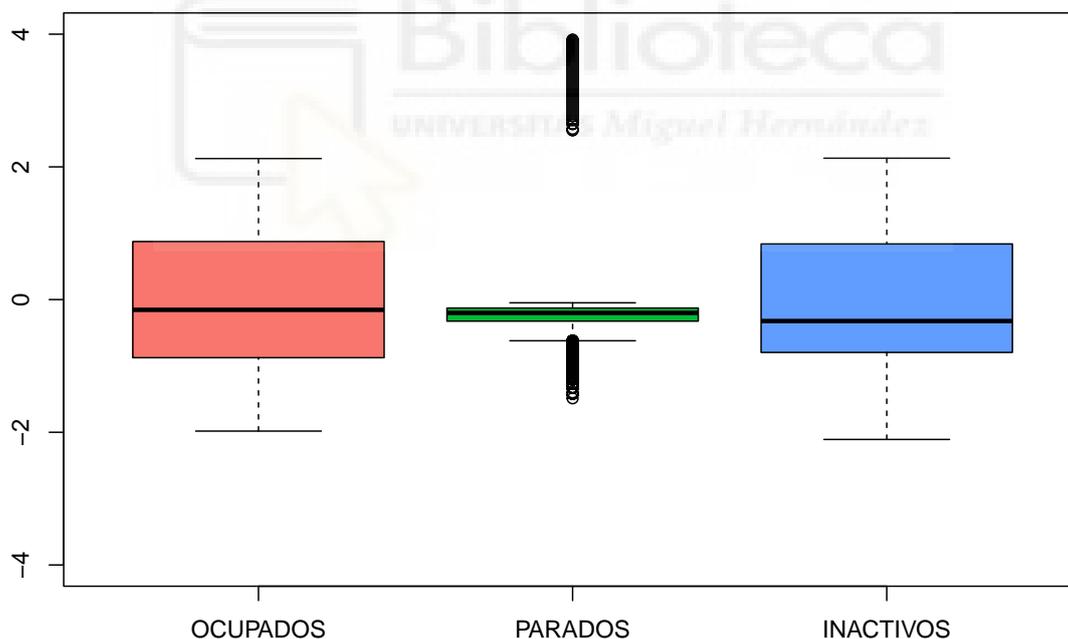


**Figura 5:** Histograma de frecuencias de residuos brutos divididos en tres categorías. Con los datos obtenidos del modelo multinomial ajustado.

Podemos observar que para las tres categorías el histograma de frecuencias de los residuos brutos muestran una forma aproximadamente simétrica y centrada en torno al cero. Esto sugiere que los residuos siguen una distribución cercana a la normalidad, lo cual es un indicador positivo de la validez del modelo ajustado.

Por último, estudiaremos las diferencias en los patrones de residuos entre las distintas categorías y determinar si existen valores atípicos o desviaciones significativas que puedan afectar el modelo ajustado.

La **Figura 6** muestra un diagrama de cajas y bigotes (también conocido como boxplot), con los residuos tipificados para cada categoría (ocupados, parados e inactivos), donde los valores atípicos se representan como puntos fuera de los bigotes, mientras que las cajas representan el rango intercuartílico y la línea central indica la mediana.



**Figura 6:** Diagrama de cajas y bigotes de los residuos tipificados por categorías. Con los datos obtenidos del modelo multinomial ajustado.

Se puede observar que en las categorías de ocupados y de inactivos, el diagrama de cajas y bigotes muestra una distribución bastante simétrica de los residuos, sin ningún valor atípico. Esto muestra que el modelo ajustado es capaz de explicar adecuadamente la variabilidad de los datos en estas dos categorías.

Sin embargo, para la categoría de parados, el diagrama de cajas y bigotes revela una mayor dispersión de los residuos y la presencia de bastantes valores atípicos por encima del rango intercuartílico. Estos valores atípicos pueden indicar observaciones inusuales o datos que no se ajustan bien al modelo.



### 5.3. Función multinom de R

El paquete `nnet` de R es una biblioteca especializada que proporciona una amplia gama de herramientas para el modelado de redes neuronales artificiales y otros tipos de modelos relacionados. Una de las funciones clave en este paquete es la función `multinom`, que se utiliza para ajustar modelos multinomiales utilizando el método de estimación de máxima verosimilitud.

La función `multinom` implementa un enfoque basado en redes neuronales para estimar los parámetros de un modelo multinomial. Utiliza el algoritmo de descenso por gradiente para maximizar la función de verosimilitud, que es una medida de la probabilidad de observar los datos dado un conjunto de parámetros.

Al llamar a la función `multinom` en R, debes proporcionar los datos en forma de una tabla de contingencia o en un formato de matriz, donde cada fila representa una observación y las columnas representan las diferentes categorías de la variable dependiente. Además, debes especificar la fórmula del modelo, que indica la relación entre la variable dependiente y las variables independientes.

El método de estimación de máxima verosimilitud busca encontrar los valores de los parámetros del modelo que maximizan la probabilidad de observar los datos que tenemos. Estos parámetros estimados representan las probabilidades de pertenecer a cada una de las categorías de la variable dependiente, dadas las variables independientes.

Una vez ajustado el modelo multinomial utilizando la función `multinom`, obtendremos un objeto que contiene información sobre los parámetros estimados, los valores ajustados y los residuos del modelo, entre otros resultados.

En resumen, el paquete `nnet` y la función `multinom` de R proporcionan una herramienta poderosa para ajustar modelos multinomiales utilizando el método de estimación de máxima verosimilitud. Estos recursos son ampliamente utilizados en la modelización estadística y son especialmente útiles cuando se trabaja con variables categóricas con más de dos categorías posibles.

## 6. Estimadores aplicados a la EPA

En esta sección, abordaremos la importancia de emplear estimadores de dominio y presentaremos la formulación de los estimadores empleados en nuestro estudio. Utilizaremos tanto estimadores directos como estimadores obtenidos a través del modelo multinomial.

Los estimadores directos son ampliamente utilizados cuando se desea estimar un parámetro poblacional basado en una muestra representativa. Estos estimadores se obtienen a través de fórmulas o cálculos específicos que consideran únicamente la información observada en la muestra. Su ventaja radica en su simplicidad y facilidad de aplicación, lo que los convierte en una opción atractiva en muchas situaciones.

Por otro lado, también utilizaremos los estimadores obtenidos en el modelo multinomial, los cuales se emplean en contextos donde los datos se distribuyen en múltiples categorías o clases. Estos estimadores se fundamentan en la teoría de la distribución multinomial, la cual modela la probabilidad conjunta de observar diferentes categorías en una muestra. A través de este enfoque, es posible estimar los parámetros del modelo y obtener información detallada sobre la distribución de las categorías en la población.

En nuestro estudio, disponemos de una muestra de la población española conocida como EPA, la cual utilizaremos para estimar parámetros en diversas subpoblaciones que denominamos dominios. Estos dominios suelen definirse como grupos socioeconómicos o áreas geográficas. En nuestro caso, hemos cruzado las provincias con el sexo para delimitar los dominios. Al hacer ese cruce, nos quedaremos con 104 dominios.

Concretamente, los indicadores que deseamos estimar son: la proporción de parados, la proporción de ocupados y la tasa de paro, todos desagregados por provincias y sexo.

Con el uso de estos estimadores, podremos obtener estimaciones precisas y representativas de los indicadores de interés en cada dominio, lo cual nos permitirá tener una visión más detallada y completa de la situación laboral en España.

A continuación, presentaremos la notación empleada para el cálculo de los indicadores:

- **Índices:** En cuanto a los índices utilizaremos la siguiente notación:
  - $d = 1, \dots, 104$  para las áreas (provincias cruzadas con sexo), donde utilizaremos la siguiente notación:
    - Los índices del 1 al 52 representarán las provincias autónomas para el grupo de hombres.
    - Los índices del 53 al 104 tendrán el mismo propósito, pero para el grupo de mujeres.
  - $j = 1, \dots, N$  para los individuos.
- **Tamaños:** En cuanto a los tamaños de la población y la muestra, utilizaremos la siguiente notación:
  - $N$  representará el tamaño de la población total. Este valor indica el número total de individuos en la población que estamos estudiando.
  - $n$  referirá al tamaño de la muestra. Este valor indica el número de individuos que han sido seleccionados de la población para formar parte de la muestra en nuestro estudio.
- **Factores de elevación:** Los factores de elevación se denotarán como  $w_{dj}$ , donde  $j = 1, \dots, N$ . Estos factores se utilizan para corregir los pesos teóricos en función de la falta de respuesta en la muestra y, posteriormente, se calibran según las cantidades conocidas de provincia-sexo.

Con esto, se busca mitigar los sesgos y mejorar la calidad de las estimaciones obtenidas a partir de la muestra, permitiendo así realizar inferencias más fiables sobre la población en estudio.

Una vez adquirido el conocimiento sobre los estimadores de dominio, procederemos a presentar los estimadores que utilizaremos en nuestro estudio, junto con su formulación matemática.

## 6.1. Proporción de Parados

La proporción de parados la calcularemos teniendo en cuenta los factores de elevación ( $w_{dj}$ ), que permitirán ajustar las estimaciones de la muestra a la población total a partir de los datos muestrales obtenidos.

- Nuestras variables son las siguientes:

- $y_{dj} = \begin{cases} 1 & \text{si } j \text{ es parado} \\ 0 & \text{en caso contrario} \end{cases}$

- $\hat{\pi}_{dj}^{par}$  = probabilidad predicha de que  $j$  sea parado según el modelo multinomial

- Nuestros estimadores son:

$$\begin{aligned} \bar{Y}_d^{dir} &= \frac{1}{N_d} \sum_{j=1}^{N_d} y_{dj} & \bar{Y}_d^{mod} &= \frac{1}{N_d} \sum_{j=1}^{N_d} \hat{\pi}_{dj}^{par} \\ \hat{Y}_d^{dir} &= \frac{\sum_{j=1}^{n_d} w_{dj} y_{dj}}{\sum_{j=1}^{n_d} w_{dj}} & \hat{Y}_d^{mod} &= \frac{\sum_{j=1}^{n_d} w_{dj} \hat{\pi}_{dj}^{par}}{\sum_{j=1}^{n_d} w_{dj}} \end{aligned}$$

Donde  $\hat{Y}_d^{dir}$  y  $\hat{Y}_d^{mod}$  son las estimaciones de la proporción de parados según estimaciones directas y según estimaciones del modelo multinomial, respectivamente.

Estas estimaciones nos permitirán obtener una medida de la proporción de parados en el dominio  $d$  utilizando los enfoques de estimación directa y multinomial, considerando los factores de elevación ( $w_{dj}$ ) para ajustar las estimaciones a la población total.

## 6.2. Proporción de Ocupados

La proporción de ocupados la calcularemos teniendo en cuenta los factores de elevación ( $w_{dj}$ ), que permitirán ajustar las estimaciones de la muestra a la población total a partir de los datos muestrales obtenidos.

- Nuestras variables son las siguientes:

$$\bullet z_{dj} = \begin{cases} 1 & \text{si } j \text{ es ocupado} \\ 0 & \text{en caso contrario} \end{cases}$$

- $\hat{\pi}_{dj}^{ocu}$  = probabilidad predicha de que  $j$  sea ocupado según el modelo multinomial

- Nuestros estimadores son:

$$\begin{aligned} \bar{Z}_d^{dir} &= \frac{1}{N_d} \sum_{j=1}^{N_d} z_{dj} & \bar{Z}_d^{mod} &= \frac{1}{N_d} \sum_{j=1}^{N_d} \hat{\pi}_{dj}^{ocu} \\ \hat{\bar{Z}}_d^{dir} &= \frac{\sum_{j=1}^{n_d} w_{dj} z_{dj}}{\sum_{j=1}^{n_d} w_{dj}} & \hat{\bar{Z}}_d^{mod} &= \frac{\sum_{j=1}^{n_d} w_{dj} \hat{\pi}_{dj}^{ocu}}{\sum_{j=1}^{n_d} w_{dj}} \end{aligned}$$

Donde  $\hat{\bar{Z}}_d^{dir}$  y  $\hat{\bar{Z}}_d^{mod}$  son las estimaciones de la proporción de ocupados según estimaciones directas y según estimaciones del modelo multinomial, respectivamente.

Estas estimaciones nos permitirán obtener una medida de la proporción de ocupados en el dominio  $d$  utilizando los enfoques de estimación directa y multinomial, considerando los factores de elevación ( $w_{dj}$ ) para ajustar las estimaciones a la población total.

### 6.3. Tasa de Paro

Las tasas de paro por provincia y sexo las calcularemos utilizando las proporciones de parados y ocupados estimados anteriormente. A partir de las estimaciones de las proporciones de parados y ocupados en cada dominio, podemos calcular las tasas de paro como el cociente entre la proporción de parados y la suma de la proporción de parados y la proporción de ocupados.

De esta manera, obtendremos las tasas de paro por provincia y sexo, que nos proporcionarán una medida de la proporción de personas desempleadas en relación a la población activa en cada dominio.

$$R_d^{dir} = \frac{\bar{Y}_d^{dir}}{\bar{Y}_d^{dir} + \bar{Z}_d^{dir}} \quad (4)$$

$$R_d^{mod} = \frac{\bar{Y}_d^{mod}}{\bar{Y}_d^{mod} + \bar{Z}_d^{mod}} \quad (5)$$

Donde  $R_d^{dir}$  y  $R_d^{mod}$  son las estimaciones de las tasas de paro según estimaciones directas y según estimaciones del modelo multinomial, respectivamente.

## 7. Resultados estimadores aplicados a la EPA

En esta sección, procederemos a mostrar detalladamente los estimadores que se obtuvieron mediante las estimaciones directas y las estimaciones obtenidas utilizando el modelo multinomial. Todas las estimaciones fueron obtenidas mediante el programa informático R.

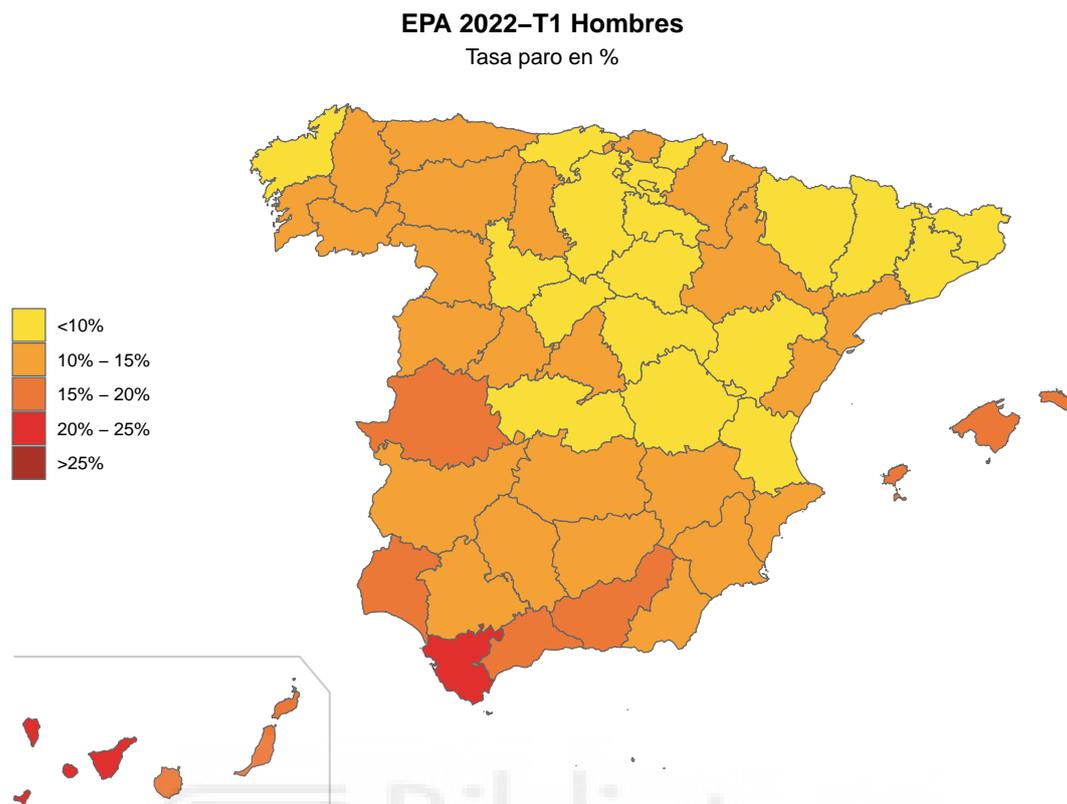
Para lograrlo, utilizaremos representaciones gráficas en forma de mapas, los cuales visualizarán de manera clara y concisa las tasas de paro correspondientes al conjunto de provincias, y dichas tasas estarán desglosadas según el sexo.

### 7.1. Estimaciones directas

En primer lugar, procederemos a representar de manera visual las tasas de paro obtenidas a través de las estimaciones directas, tal como se describe en la sección 6 del trabajo, concretamente utilizando la ecuación 4. Para una mejor comprensión, hemos dividido las tasas de paro en cinco intervalos:  $[0\% - 10\%]$ ,  $[10\% - 15\%]$ ,  $[15\% - 20\%]$ ,  $[20\% - 25\%]$  y  $> 25\%$ .

#### 7.1.1. Tasas de paro masculinas

En la **Figura 7**, se presenta un mapa de España dividido por provincias. Cada provincia está sombreada o coloreada de acuerdo con la tasa de paro para los hombres en esa región. Se utilizan diferentes colores o tonos para indicar los diferentes intervalos de tasas de paro, como se mencionó anteriormente:



**Figura 7:** Representación gráfica de las tasas de paro por provincias en hombres mediante las estimaciones directas.

Las comunidades autónomas más afectadas por el desempleo son Andalucía, Extremadura, las Islas Canarias y las ciudades autónomas de Ceuta y Melilla. En particular, Ceuta tiene la tasa de paro más alta de todo el territorio nacional, con un 24,78 %. Cádiz y Santa Cruz de Tenerife también superan el 20 % de paro.

En contraste, Huesca es la provincia con la tasa de paro más baja, con un 3,46 %, seguida de cerca por Gipuzcoa y Lérida, todas ellas por debajo del 6 %.

La media de las tasas de paro en hombres de todas las regiones incluidas en el estudio es del 12,09 % y presenta una variabilidad bastante grande entre las distintas regiones que componen España como hemos visto anteriormente.

La media de las tasas de paro para los hombres en todas las regiones incluidas en el estudio es del 12,09 %. Sin embargo, se puede observar que existe una gran variabilidad entre las distintas regiones que conforman el territorio español, donde en las provincias

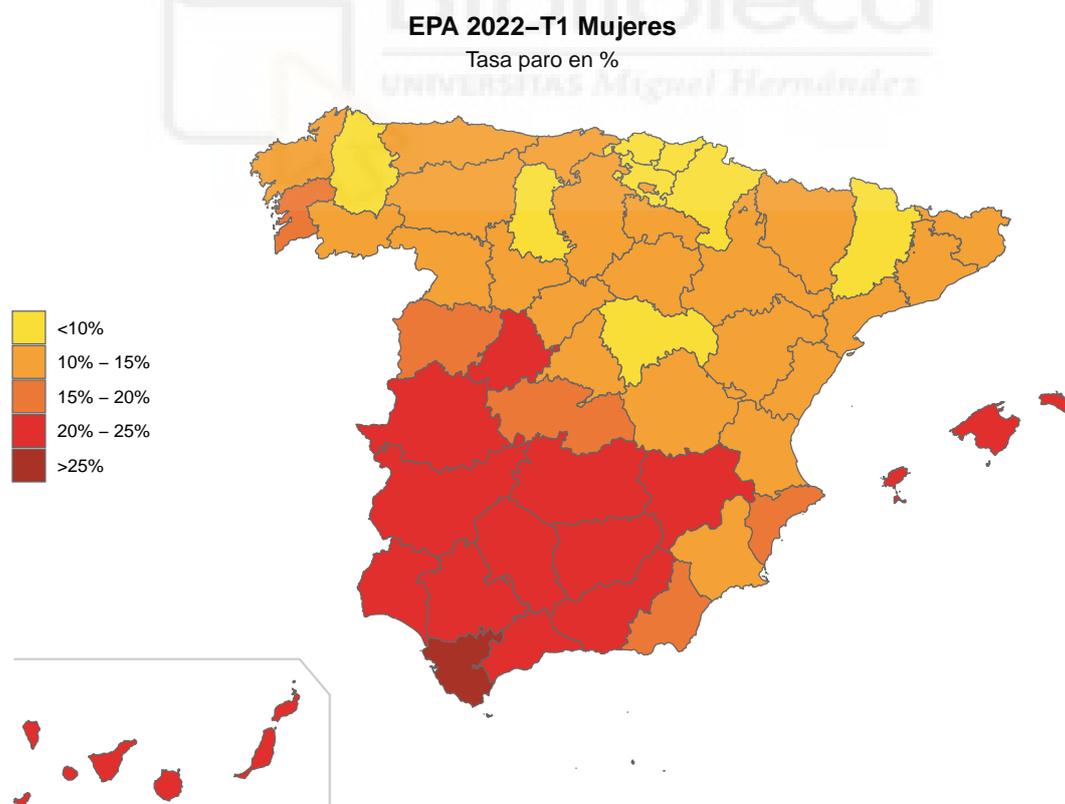
situadas en el extremo norte del país presentan una tasa de desempleo notablemente inferior en comparación con las provincias ubicadas en el extremo sur.

Estos datos resaltan las disparidades significativas en las tasas de paro a nivel regional y demuestran la importancia de comprender y abordar las causas subyacentes del desempleo en cada región de España.

### 7.1.2. Tasas de paro femeninas

A continuación, procederemos a representar visualmente las tasas de desempleo correspondientes al género femenino, basándonos en las estimaciones directas realizadas.

En la **Figura 8**, se presenta un mapa de España dividido por provincias. Cada provincia está sombreada o coloreada de acuerdo con la tasa de paro para las mujeres en esa región. Se utilizan diferentes colores o tonos para indicar los diferentes intervalos de tasas de paro, como se mencionó anteriormente:



**Figura 8:** Representación gráfica de las tasas de paro por provincias en mujeres mediante las estimaciones directas.

Al igual que en el caso de los hombres, las comunidades autónomas más afectadas por el desempleo en mujeres son aquellas situadas en el sur de España, como Andalucía, Extremadura, las Islas Canarias y las ciudades autónomas de Ceuta y Melilla. Específicamente, la tasa de desempleo en Ceuta es la más alta de todo el territorio nacional, alcanzando un 36,42 %.

Por otro lado, en las comunidades autónomas del norte de España, como Galicia, Asturias, el País Vasco y Navarra, se observan las tasas de paro más bajas. En particular, la provincia de Vizcaya tiene la menor tasa de paro para las mujeres, con un 7.16 %, seguida de cerca por Álava, Palencia y Lérida, todas ellas con tasas por debajo del 9.00 %.

La media de las tasas de paro para las mujeres en todas las regiones incluidas en el estudio es del 15.60 %. Sin embargo, existe una variabilidad significativa entre las distintas regiones que componen España, como se ha mencionado anteriormente en el caso de los hombres.

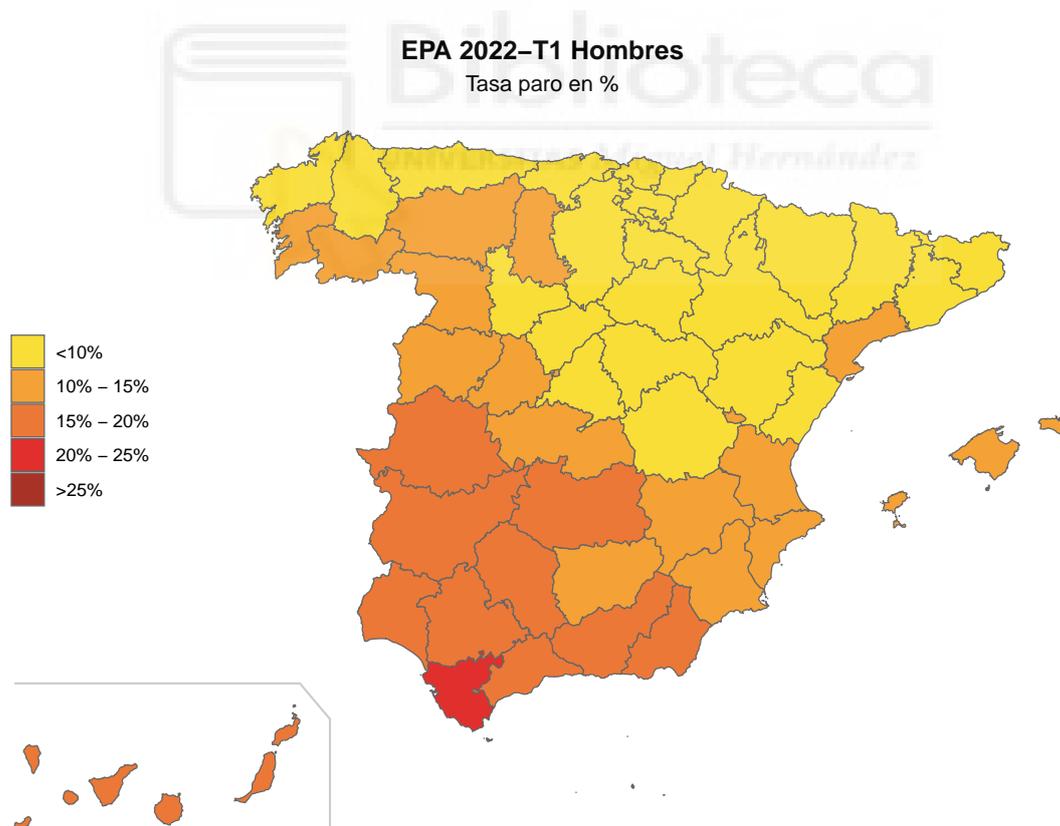
Cabe destacar, en términos generales, que las tasas de paro para las mujeres suelen ser significativamente más altas que las tasas de paro para los hombres.

## 7.2. Estimaciones modelo multinomial

En segundo lugar, procederemos a representar de manera visual las tasas de paro obtenidas a través de las estimaciones obtenidas utilizando el modelo multinomial, tal como se describe en la sección 6 del trabajo, concretamente utilizando la ecuación 5. Para una mejor comprensión, hemos dividido las tasas de paro en cinco intervalos:  $[0\% - 10\%]$ ,  $[10\% - 15\%]$ ,  $[15\% - 20\%]$ ,  $[20\% - 25\%]$  y  $> 25\%$ .

### 7.2.1. Tasas de paro masculinas

En la **Figura 9**, se presenta un mapa de España dividido por provincias. Cada provincia está sombreada o coloreada de acuerdo con la tasa de paro para los hombres en esa región. Se utilizan diferentes colores o tonos para indicar los diferentes intervalos de tasas de paro, como se mencionó anteriormente:



**Figura 9:** Representación gráfica de las tasas de paro por provincias en hombres mediante las estimaciones del modelo multinomial.

Las comunidades más afectadas por el desempleo son Andalucía, Extremadura, las Islas Canarias y las ciudades autónomas de Ceuta y Melilla. En particular, Ceuta tiene la tasa de paro más alta de todo el territorio nacional, con un 25.76 %. Cádiz y Melilla también superan el 20 % de paro.

En contraste, Gipúzcoa es la provincia con la tasa de paro más baja, con un 5.44 %, seguida de cerca por Lérida, Huesca o Soria, todas ellas por debajo del 7 %.

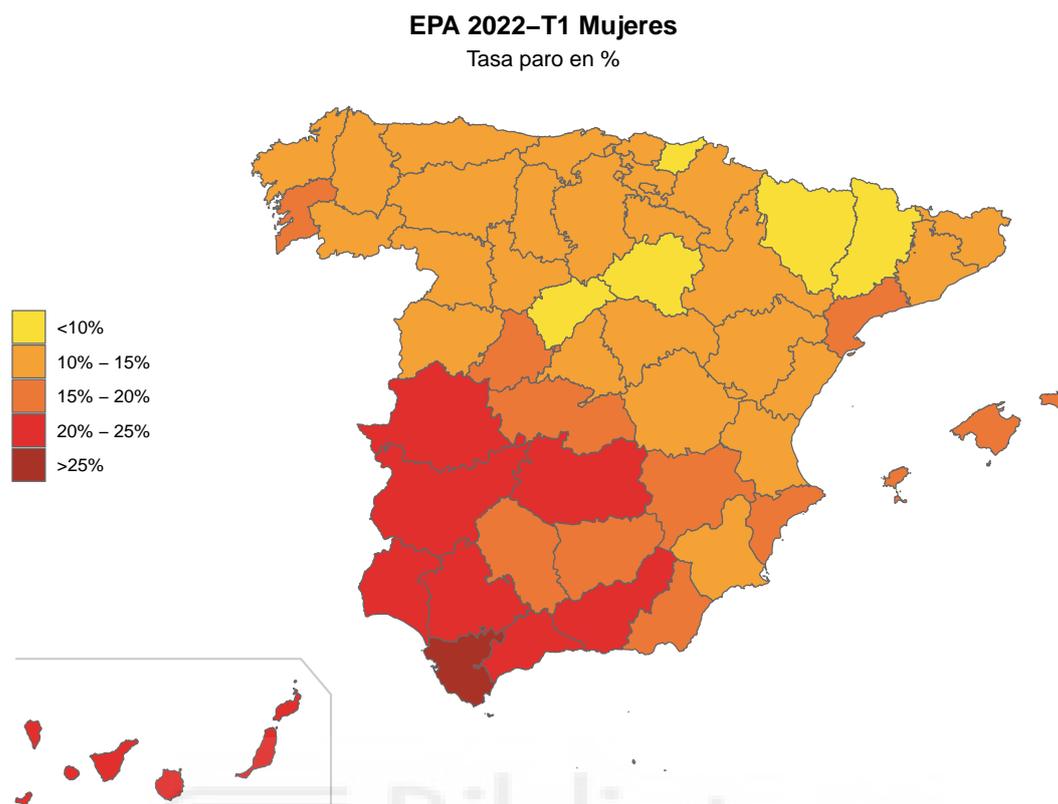
La media de las tasas de paro para los hombres en todas las regiones incluidas en el estudio es del 11.53 %. Sin embargo, se puede observar que existe una gran variabilidad entre las distintas regiones que componen el territorio español, donde en las provincias situadas en el extremo norte del país presentan una tasa de paro notablemente inferior en comparación con las provincias ubicadas en el extremo sur.

Estos datos resaltan las disparidades significativas en las tasas de paro a nivel regional y demuestran la importancia de comprender y abordar las causas subyacentes del desempleo en cada región de España.

### 7.2.2. Tasas de paro femeninas

A continuación, procederemos a representar visualmente las tasas de desempleo correspondientes al género femenino, basándonos en las estimaciones obtenidas utilizando el modelo multinomial.

En la **Figura 10**, se presenta un mapa de España dividido por provincias. Cada provincia está sombreada o coloreada de acuerdo con la tasa de paro para las mujeres en esa región. Se utilizan diferentes colores o tonos para indicar los diferentes intervalos de tasas de paro, como se mencionó anteriormente:



**Figura 10:** Representación gráfica de las tasas de paro por provincias en mujeres mediante las estimaciones del modelo multinomial.

Al igual que en el caso de los hombres, las comunidades autónomas más afectadas por el desempleo en mujeres son aquellas situadas en el sur de España, como Andalucía, Extremadura, las Islas Canarias y las ciudades autónomas de Ceuta y Melilla. Específicamente, la tasa de desempleo en Ceuta es la más alta de todo el territorio nacional, alcanzando un 32,23%.

Por otro lado, en las comunidades autónomas del norte de España, como Galicia, Asturias, el País Vasco y Navarra, se observan las tasas de paro más bajas. En particular, la provincia de Lérida tiene la menor tasa de paro para las mujeres, con un 7.53%, seguida de cerca por Huesca, Gipúzcoa y Soria, todas ellas con tasas por debajo del 9.00%.

La media de las tasas de paro para las mujeres en todas las regiones incluidas en el estudio es del 15.21%. Sin embargo, existe una variabilidad significativa entre las distintas regiones que componen España, como se ha mencionado anteriormente en el caso de los hombres.

Cabe destacar, en términos generales, que las tasas de paro para las mujeres suelen ser significativamente más altas que las tasas de paro para los hombres.



## 8. Conclusiones

A continuación, presentaremos las conclusiones derivadas de los capítulos anteriores. En primer lugar, abordaremos las conclusiones de carácter general, seguidas de las conclusiones específicas obtenidas para cada caso analizado.

En términos generales, podemos observar que no existen diferencias geográficas relevantes entre los sexos en cuanto a la tasa de paro. Esto implica que si una provincia muestra un alto nivel de desempleo para los hombres, es probable que también lo presente para las mujeres, y viceversa. En el Apéndice, se encuentra disponible la **Tabla 3**, que muestra las tasas de paro desglosadas por sexo y provincias, obtenidas a partir de las estimaciones realizadas en los capítulos anteriores.

Si analizamos las tasas de paro en términos relativos, podemos observar claramente un patrón geográfico al que podríamos denominar “**factor meridional**”. Las regiones del norte del país muestran tasas de paro significativamente más bajas en comparación con otras áreas, especialmente en contraste con las zonas del sur, donde se registran las tasas de paro más altas.

Este patrón geográfico revela una disparidad notable en cuanto a la situación del mercado laboral entre el norte y el sur del país. Las regiones del norte, como País Vasco, Navarra o Cantabria, presentan tasas de paro relativamente bajas, lo que sugiere una mayor estabilidad y oportunidades de empleo en comparación con otras áreas.

Por otro lado, las zonas del sur de España, como Andalucía, Extremadura, Ceuta, Melilla o las Islas Canarias, muestran tasas de paro más elevadas. Esto indica que estas regiones enfrentan desafíos significativos en términos de generación de empleo y oportunidades laborales para su población.

Es importante tener en cuenta esta disparidad regional al analizar la situación del mercado laboral en España, ya que nos permite comprender que existen diferencias significativas en las oportunidades de empleo y las condiciones laborales según la ubicación geográfica.

Este “factor meridional” puede estar influenciado por una serie de factores, como la estructura económica regional, la disponibilidad de recursos, las políticas laborales y de desarrollo, entre otros. Comprender estas dinámicas regionales es fundamental para diseñar estrategias y políticas que promuevan la creación de empleo y reduzcan las tasas de paro en las áreas más afectadas.

En resumen, el análisis de las tasas de paro en términos relativos revela un patrón geográfico en España, donde las regiones del norte presentan tasas de paro más bajas en comparación con las áreas del sur. Esta disparidad regional destaca la importancia de adoptar enfoques específicos para abordar los desafíos del mercado laboral en diferentes partes del país.

Cabe resaltar, además, que **las mujeres experimentan tasas de paro más altas en comparación con los hombres**, y este fenómeno es evidente al observar el mapa geográfico que hemos identificado como el “factor meridional”.

Este hallazgo refleja una disparidad de género en el mercado laboral, donde las mujeres enfrentan mayores dificultades para encontrar empleo en varias regiones del país, especialmente en aquellas áreas donde las tasas de paro son más altas. Esta situación puede estar relacionada con una serie de factores, como la segregación ocupacional, la discriminación de género, las brechas salariales y la falta de oportunidades laborales equitativas.

La identificación de estas diferencias de género en las tasas de paro resalta la importancia de promover políticas y acciones que fomenten la igualdad de oportunidades en el mercado laboral. Es fundamental abordar las barreras y desafíos específicos que enfrentan las mujeres, como el acceso a la formación y capacitación laboral, la conciliación entre el trabajo y la vida personal, y la eliminación de la discriminación y los estereotipos de género en el ámbito laboral.

La equidad de género en el empleo no solo es esencial para promover la justicia social, sino que también contribuye al desarrollo económico sostenible de un país. La participación plena y activa de las mujeres en el mercado laboral no solo impulsa el crecimiento económico, sino que también tiene efectos positivos en la reducción de la

pobreza, el aumento de la productividad y el fortalecimiento de la cohesión social.

En conclusión, la disparidad de género en las tasas de paro, evidenciada en el “factor meridional” geográfico, destaca la necesidad de implementar políticas inclusivas y acciones concretas para eliminar las barreras que enfrentan las mujeres en el mercado laboral y promover la igualdad de oportunidades para todos los individuos, independientemente de su género.

En resumen, es importante considerar que las tasas de paro son altas en general. La existencia de tasas de paro superiores al 20 % e incluso al 30 % es motivo de preocupación para la economía de un país, sin importar las causas que hayan contribuido a estas cifras.

Si bien es cierto que algunas provincias pueden presentar tasas de paro más bajas, a nivel global y en comparación con otros países de la Unión Europea, las tasas en España son considerablemente altas. Mientras que la tasa de paro para hombres en la UE es del 6.70 % y para mujeres es del 7.40 %, en España estas cifras ascienden al 12.09 % para hombres y al 15.60 % para mujeres, es decir, casi el doble. [7]

Estas diferencias en las tasas de paro entre España y otros países de la UE indican que existen desafíos significativos en el mercado laboral español. Estos desafíos pueden estar relacionados con la falta de oportunidades de empleo, la baja inversión en sectores productivos, la precariedad laboral y otros factores socioeconómicos.

En consecuencia, es necesario adoptar medidas y políticas eficaces para reducir las tasas de paro y promover un mercado laboral más dinámico y equitativo. Esto implica impulsar la creación de empleo, fomentar la inversión en sectores estratégicos, mejorar la formación y capacitación de los trabajadores, así como garantizar condiciones laborales justas y dignas.

En última instancia, la reducción del desempleo y la mejora de las condiciones laborales no solo beneficiarán a los individuos y las familias, sino que también fortalecerán la economía en su conjunto, impulsando el crecimiento sostenible y la prosperidad para todos.

## 9. Apéndice A - Tabla de coeficientes

**Tabla 2:** Tabla donde se presentan las estimaciones de los parámetros (coeff), los errores estándar (s.e.) y los intervalos de confianza al 95 %.

	PARADOS				INACTIVOS			
	coeff	s.e.	2.5 %	97.5 %	coeff	s.e.	2.5 %	97.5 %
(Intercept)	-0.56	0.15	-0.86	-0.27	0.14	0.08	-0.01	0.30
EDAD1	-0.02	0.00	-0.02	-0.01	0.03	0.00	0.03	0.03
SEXOn6	0.39	0.02	0.34	0.44	0.45	0.01	0.42	0.47
FORMAn2	-0.94	0.05	-1.03	-0.85	-1.71	0.03	-1.76	-1.66
FORMAn3	-1.74	0.05	-1.84	-1.64	-2.91	0.03	-2.97	-2.86
PROVn2	0.43	0.18	0.09	0.78	-0.16	0.10	-0.35	0.03
PROVn3	0.38	0.16	0.06	0.70	-0.14	0.09	-0.31	0.03
PROVn4	0.37	0.17	0.03	0.70	-0.41	0.10	-0.60	-0.23
PROVn5	0.36	0.19	-0.02	0.73	-0.06	0.10	-0.26	0.14
PROVn6	0.67	0.16	0.37	0.98	-0.15	0.09	-0.31	0.02
PROVn7	0.44	0.15	0.14	0.75	-0.17	0.08	-0.33	-0.01
PROVn8	0.02	0.15	-0.28	0.32	-0.18	0.08	-0.33	-0.03
PROVn9	-0.07	0.19	-0.45	0.30	-0.23	0.10	-0.42	-0.04
PROVn10	0.66	0.16	0.34	0.98	0.01	0.09	-0.17	0.18
PROVn11	1.10	0.15	0.80	1.41	0.18	0.09	0.01	0.35
PROVn12	0.05	0.17	-0.28	0.39	-0.05	0.09	-0.22	0.12
PROVn13	0.61	0.16	0.30	0.93	-0.05	0.09	-0.22	0.12
PROVn14	0.58	0.16	0.26	0.90	-0.11	0.09	-0.28	0.07
PROVn15	0.00	0.15	-0.30	0.31	0.06	0.08	-0.09	0.21
PROVn16	-0.19	0.20	-0.57	0.20	-0.27	0.10	-0.46	-0.08
PROVn17	-0.08	0.18	-0.43	0.27	-0.27	0.09	-0.44	-0.09
PROVn18	0.76	0.16	0.45	1.07	0.10	0.09	-0.07	0.27
PROVn19	-0.13	0.19	-0.50	0.25	-0.23	0.10	-0.42	-0.04
PROVn20	-0.32	0.21	-0.73	0.08	0.16	0.09	-0.02	0.35
PROVn21	0.58	0.18	0.24	0.93	-0.06	0.10	-0.25	0.14

PROVn22	-0.45	0.23	-0.90	-0.00	-0.13	0.10	-0.33	0.07
PROVn23	0.48	0.16	0.16	0.80	-0.03	0.09	-0.21	0.14
PROVn24	0.27	0.17	-0.07	0.61	0.02	0.09	-0.15	0.20
PROVn25	-0.51	0.22	-0.95	-0.07	-0.23	0.10	-0.42	-0.03
PROVn26	0.09	0.17	-0.24	0.41	-0.01	0.08	-0.18	0.15
PROVn27	0.06	0.17	-0.29	0.40	0.02	0.09	-0.15	0.19
PROVn28	0.17	0.15	-0.11	0.46	-0.16	0.07	-0.30	-0.01
PROVn29	0.68	0.16	0.37	0.99	0.01	0.09	-0.16	0.18
PROVn30	0.19	0.15	-0.11	0.50	-0.18	0.08	-0.34	-0.03
PROVn31	-0.06	0.17	-0.39	0.26	-0.04	0.08	-0.21	0.12
PROVn32	0.18	0.18	-0.16	0.53	0.26	0.09	0.08	0.42
PROVn33	0.21	0.16	-0.10	0.53	0.34	0.08	0.19	0.50
PROVn34	0.14	0.20	-0.24	0.52	-0.08	0.10	-0.28	0.12
PROVn35	0.74	0.15	0.43	1.04	-0.11	0.08	-0.27	0.06
PROVn36	0.35	0.15	0.05	0.64	0.09	0.08	-0.06	0.24
PROVn37	0.26	0.18	-0.09	0.62	-0.03	0.09	-0.22	0.15
PROVn38	0.77	0.15	0.46	1.07	-0.06	0.08	-0.23	0.10
PROVn39	0.14	0.16	-0.18	0.46	0.20	0.08	0.04	0.36
PROVn40	-0.25	0.21	-0.66	0.16	-0.25	0.10	-0.44	-0.05
PROVn41	0.68	0.15	0.38	0.97	0.11	0.08	-0.04	0.26
PROVn42	-0.30	0.21	-0.71	0.10	-0.25	0.10	-0.45	-0.06
PROVn43	0.32	0.17	-0.02	0.66	-0.04	0.09	-0.22	0.14
PROVn44	-0.15	0.20	-0.55	0.25	-0.18	0.10	-0.37	0.01
PROVn45	0.27	0.16	-0.05	0.59	-0.32	0.09	-0.48	-0.14
PROVn46	0.17	0.15	-0.12	0.48	-0.06	0.08	-0.22	0.09
PROVn47	0.00	0.18	-0.35	0.35	0.16	0.09	-0.01	0.32
PROVn48	0.10	0.17	-0.24	0.43	0.04	0.08	-0.12	0.21
PROVn49	0.31	0.20	-0.09	0.70	0.05	0.10	-0.15	0.26
PROVn50	0.03	0.16	-0.29	0.35	-0.21	0.08	-0.37	-0.05
PROVn51	1.15	0.19	0.77	1.53	-0.20	0.13	-0.46	0.06
PROVn52	0.90	0.21	0.49	1.31	-0.12	0.14	-0.40	0.17

## 10. Apéndice B - Tabla de resultados

**Tabla 3:** Tasas de paro por provincia y sexo (en porcentaje) obtenidas mediante estimaciones directas y estimaciones del modelo multinomial.

#	Provincia	Estimaciones directas		Estimaciones modelo	
		Hombre	Mujer	Hombre	Mujer
01	Álava	9.99	7.95	7.81	10.56
02	Albacete	11.42	22.59	14.16	18.13
03	Alicante	14.69	15.55	12.38	16.57
04	Almería	14.77	17.86	15.45	18.71
05	Ávila	10.62	20.83	12.83	16.21
06	Badajoz	14.63	24.24	16.79	21.82
07	Baleares	17.17	20.60	13.55	17.52
08	Barcelona	9.63	11.04	8.66	11.67
09	Burgos	7.36	13.18	7.70	10.29
10	Cáceres	16.95	21.66	16.72	21.01
11	Cádiz	21.42	31.51	23.39	29.35
12	Castellón	10.01	12.88	9.16	12.50
13	Ciudad Real	13.65	24.65	15.94	20.70
14	Córdoba	13.15	20.70	15.81	19.90
15	La Coruña	8.79	11.51	8.11	10.78
16	Cuenca	8.58	12.01	8.26	10.91
17	Gerona	9.90	10.26	8.34	11.06
18	Granada	18.37	20.49	17.28	22.28
19	Guadalajara	9.54	9.72	8.05	11.45
20	Guipúzcoa	4.76	9.63	5.44	7.85
21	Huelva	17.15	20.87	16.39	20.60
22	Huesca	3.46	10.48	6.02	7.58
23	Jaén	13.70	24.45	14.60	19.02

---

24	León	11.74	14.01	11.04	14.31
25	Lérida	5.91	8.84	5.63	7.53
26	La Rioja	9.89	14.03	9.10	12.16
27	Lugo	10.11	9.27	8.71	11.57
28	Madrid	10.45	13.85	9.44	12.94
29	Málaga	16.75	24.06	16.38	21.45
30	Murcia	11.39	15.00	11.53	14.72
31	Navarra	11.48	8.97	7.76	10.54
32	Orense	12.19	11.09	10.04	12.91
33	Asturias	10.06	14.02	9.32	12.89
34	Palencia	12.05	8.34	10.14	13.19
35	Las Palmas	18.77	22.88	17.32	22.20
36	Pontevedra	10.72	16.47	11.56	15.21
37	Salamanca	11.22	15.25	10.88	14.02
38	Santa Cruz de Tenerife	21.18	20.05	17.08	21.91
39	Cantabria	8.45	13.17	9.13	12.55
40	Segovia	6.07	11.91	7.32	9.35
41	Sevilla	13.34	22.96	15.24	20.23
42	Soria	6.70	10.45	6.94	9.22
43	Tarragona	12.40	13.78	11.94	15.49
44	Teruel	8.47	12.82	7.81	10.44
45	Toledo	8.72	16.91	12.24	15.91
46	Valencia	9.93	13.50	10.05	13.53
47	Valladolid	8.21	10.83	8.09	11.00
48	Vizcaya	10.36	7.17	8.09	11.19
49	Zamora	11.67	13.93	11.43	14.86
50	Zaragoza	10.24	12.63	8.76	11.63
51	Ceuta	24.78	36.42	25.76	32.23
52	Melilla	19.19	28.12	20.51	26.51

---

## 11. Apéndice C - Código R

```
# Cargamos las librerías:
library(readr) #lectura del archivo
library(tidyverse) #manejar datos
library(nnet) #modelo multinomial
library(mapSpain) #mapa de España
library(rgdal) #mapa de España

# Lectura de datos:
datos<-read_csv2("EPA_2022T1.csv")

attach(datos)
dim(datos)

# Generar la matriz de variables objetivo:
OCUPADOS<-as.numeric(AOI==3 | AOI==4)
PARADOS<-as.numeric(AOI==5 | AOI ==6)
INACTIVOS<-as.numeric(AOI==7 | AOI==8 | AOI==9)
Y<-cbind(OCUPADOS, PARADOS, INACTIVOS)

# Transformamos y dividimos la variable NFORMA en 3 niveles:
N1<-as.numeric(NFORMA == "AN" | NFORMA == "P1" | NFORMA == "P2")
N2<-as.numeric(NFORMA == "S1" | NFORMA == "SG" | NFORMA == "SP")
N3<-as.numeric(NFORMA == "SU" )
FORMA<- 1 * N1 + 2 * N2 + 3 * N3
head(FORMA,15)

# Ponemos las variables que son factores como factores:
FORMAn<-factor(FORMA)
```

```
SEX0n<-factor(SEX01)
PROVn<-factor(PROV)

# Modelo multinomial:
modymax<-multinom(Y ~ PROVn + EDAD1 + SEX0n + FORMAn , data=datos)
summary(modymax)

# Coeficientes modelo modymax:
coeficientes<-coef(modymax)[1,]
tabla_coef<-data.frame(Coeficientes = coeficientes)

# Valores ajustados:
valores_ajust<-predict(modymax, type = "probs")

#### ESTIMACIONES DIRECTAS ####

# Cogemos ahora solo las filas sin NAs en AOI:
datos1<-datos[complete.cases(datos$AOI),]

# Parados:
y_dj<-as.numeric(datos1$AOI == 5 | datos1$AOI == 6)

# Ocupados:
z_dj<-as.numeric(datos1$AOI == 3 | datos1$AOI == 4)

datos1<-cbind(datos1, y_dj, z_dj)

# Proporción de parados ponderada con el FACTOREL
datos_parados1<-datos1 %>%
  group_by(PROV,SEX01) %>%
```

```

summarise(prop.parados=sum(FACTOREL*y_dj)/sum(FACTOREL)*100)

# Proporción de ocupados ponderada con el FACTOREL
datos_ocupados1<-datos1 %>% group_by(PROV,SEX01) %>%
  summarise(prop.ocupados=sum(FACTOREL*z_dj)/sum(FACTOREL)*100)

# Tasa de paro:
tasa_paro1<-c()
for (i in 1:104) {
  tasa_paro1[i]<-datos_parados1$prop.parados[i] /
  (datos_parados1$prop.parados[i] + datos_ocupados1$prop.ocupados[i])
}

# Datos completos:
datos_nombres <- datos1 %>% group_by(PROV,SEX01) %>%
  summarise()
datos_comp1<-cbind(datos_nombres, tasa_paro_muestral = tasa_paro1*100)

#### ESTIMACIONES MODELO MULTINOMIAL ####

#Cogemos ahora solo las filas sin NAs en AOI:
datos2<-datos[complete.cases(datos$AOI),]
datos2<-cbind(datos2, OCUPADOS = valores_ajust[,1],
              PARADOS = valores_ajust[,2])

# Proporción de parados:
datos_parados2 <- datos2 %>% group_by(PROV,SEX01) %>%
  summarise(prop.parados = sum(FACTOREL*PARADOS , na.rm = TRUE) /
            sum(FACTOREL) * 100)

```

```
# Proporción de ocupados:
datos_ocupados2 <- datos2 %>% group_by(PROV,SEX01) %>%
  summarise(prop.ocupados = sum(FACTOREL * OCUPADOS, na.rm = TRUE) /
  sum(FACTOREL) * 100)

# Tasa de paro:
tasa_paro2<-c()
for (i in 1:104) {
  tasa_paro2[i]<-datos_parados2$prop.parados[i] /
  (datos_parados2$prop.parados[i] + datos_ocupados2$prop.ocupados[i])
}

# Datos completos:
datos_nombres <- datos1 %>% group_by(PROV,SEX01) %>%
  summarise()
datos_comp2<-cbind(datos_nombres, tasa_paro_multi = tasa_paro2*100)
```

## 12. Bibliografía

- [1] Instituto Nacional de Estadística. *Encuesta de Población Activa (EPA). Documento Técnico*. [https://www.ine.es/epa02/documento\\_tecnico.pdf](https://www.ine.es/epa02/documento_tecnico.pdf)
- [2] Instituto Nacional de Estadística. *Encuesta de Población Activa Metodología 2021*. <https://www.ine.es/inebaseDYN/epa30308/docs/resumetepa21.pdf>
- [3] BRAGE ESCALONA, M. *Análisis de datos categóricos: regresión logística y multinomial*. Trabajo Fin de Grado. Sección de Matemáticas. Universidad de La Laguna.
- [4] JULIO MULERO Y JUAN MATÍAS SEPULCRE (2016). *L<sup>A</sup>T<sub>E</sub>X con palabras clave*. Publicacions Universitat D'Alacant.
- [5] DOMINGO MORALES GONZÁLEZ (2021). *Modelos lineales generalizados*. Estadística Avanzada. Universidad Miguel Hernández.
- [6] *MapSpain: Administrative boundaries of Spain, including CCAA, provinces and municipalities*. <https://ropenspain.github.io/mapSpain/>
- [7] *Idescat. Indicadores de la Unión Europea. Tasa de paro por sexo*. <https://www.idescat.cat/indicadors/?id=ue&n=10114&lang=es&t=202100>
- [8] BRIAN RIPLEY Y WILLIAM VENABLES (2023). *Feed-Forward Neural Networks and Multinomial Log-Linear Models* <https://cran.r-project.org/web/packages/nnet/nnet.pdf>
- [9] HADLEY WICKHAM (2016). *ggplot2: Elegant Graphics for Data Analysis* <https://ggplot2.tidyverse.org>