



FACULTAD DE CIENCIAS SOCIALES Y JURÍDICAS

GRADO EN ESTADÍSTICA EMPRESARIAL

TRABAJO FIN DE GRADO

ESTIMACIÓN DE TASAS DE PARO POR PROVINCIA Y SEXO

Presentado por: Pablo Pérez Palazón

Tutor: Prof. Dr. Domingo Morales González

Curso académico 2020/2021

TABLA DE CONTENIDOS

1-INTRODUCCIÓN	2
2-INFORME TÉCNICO EPA	4
3-DESCRIPCIÓN DEL FICHERO	13
4-ESTIMADORES DE DOMINIO.....	18
5-ESTIMADORES APLICADOS A LA EPA (T4 21).....	21
6-CONCLUSIONES.....	28
7-IMPLEMENTACIÓN DE LOS ESTIMADORES EN RSTUDIO.....	30
8-APÉNDICE A: TABLAS.....	39
9-APÉNDICE B: CUESTIONARIO EPA.....	47
9-REFERENCIAS.....	49

1-INTRODUCCIÓN

En este estudio trataremos de conocer las características de la población española en relación con el mercado de trabajo. Para esto estimaremos las tasas de paro, los totales de parados y los totales de desempleados en España por provincia y sexo. Para esto tomaremos como referencia la Encuesta de Población Activa (de ahora en adelante EPA), elaborada por el INE, del primer trimestre del año 2021.

Presentaremos los resultados de forma visual y clara haciendo uso de mapas y gráficos que ilustren los resultados obtenidos. Combinaremos la presentación de resultados con tablas donde se aporte de forma más precisa, pero no tan concisa, la información. El análisis de las estimaciones obtenidas permitirá cuantificar, y saber de primera mano, la realidad que vive el mercado laboral español.

Todo esto se presenta en distintos capítulos. En primera instancia describiremos los aspectos técnicos de la EPA. Para ello debemos conocer cómo se realiza y cuáles son las características técnicas de esta importante encuesta de cuyos datos nos serviremos para nuestras estimaciones.

Más tarde, pasaremos a estudiar los datos que vamos a tratar prestando especial atención a aquellas variables que consideramos más relevantes para el estudio. En esta parte presentaremos un breve análisis descriptivo que nos permitirá tener una visión más global de los datos que estamos tratando.

El siguiente paso sería estimar los indicadores del mercado laboral propuestos. Previamente se presenta una breve formulación matemática de todos aquellos estimadores usados. La programación e implementación de los estimadores y de todos los cálculos necesarios en este trabajo se ha realizado con el lenguaje de programación R bajo el entorno de Rstudio.

Finalmente presentamos los resultados obtenidos y damos las conclusiones derivadas de éstos. Todo seguido de un apéndice con las tablas donde se presentan todos los resultados obtenidos, donde el lector puede consultar de forma detallada los datos diferenciados por sexo y provincia.

Se emplean las variables que se consideran más relevantes sobre un total de 119. Sobre estos datos emplearemos estimadores de dominio directos que arrojarán nuestras estimaciones sobre totales de parados, totales de ocupados y tasas de paro. Todas estas estimaciones estarán desglosadas por sexo y provincia lo que nos permitirá también cuantificar la situación del mercado laboral español por género y localización geográfica.



2-INFORME TÉCNICO EPA

En esta sección presentamos un breve resumen sobre la EPA y su ficha técnica.

La EPA es una investigación por muestreo de periodicidad trimestral, dirigida a la población que reside en viviendas familiares del territorio nacional y cuya finalidad es averiguar las características de dicha población en relación con el mercado de trabajo. Esta misma encuesta empezó a realizarse en 1964 por el Instituto Nacional de Estadística.

Se trata de una encuesta de tipo continuo para conocer la actividad económica del país. Otro tipo de fuentes estadísticas como el censo o el paro registrado no son tan adecuadas para satisfacer los objetivos planteados. De este modo tenemos una encuesta que satisface la necesidad de tener una encuesta que refleje la realidad de la población activa, establecidas por la Unión Europea.

Para poder comprender la naturaleza de la EPA debemos conocer que entendemos por población económicamente activa, población ocupada, población subempleada, población parada o desempleada y población económicamente inactiva.

La Oficina Internacional de Trabajo describe a la población económicamente activa como el *conjunto de personas que, en un período de referencia dado, suministran mano de obra para la producción de bienes y servicios económicos o que están disponibles y hacen gestiones para incorporarse a dicha producción*. En nuestro caso vienen a ser las personas de 16 y más años, dado que los 16 años es el límite de escolarización obligatoria y la edad mínima legal para trabajar, que trabajan o están en disposición y tienen la voluntad de hacerlo.

Mientras que la 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. Dicha población económicamente activa está constituida por todas aquellas personas de 16 años en adelante y que cumplan con las especificaciones dadas.

Entendemos por población subempleada a aquellas personas que desean trabajar más tiempo o que directamente están disponibles para trabajar más horas, por lo que sus horas de trabajo son insuficientes.

Consideramos desempleadas a todas aquellas personas que siendo mayores de 16 años reúnan las tres siguientes características:

- No tener un empleo.
- Estar en búsqueda activa de un trabajo.
- Estas en disposición para trabajar.

Por último, consideramos población económicamente inactiva a todas las personas también mayores de 16 años que no queden recogidas en ninguna de las categorías anteriores, pudiendo ser entre otros:

- Personas que se ocupan del hogar.
- Estudiantes.
- Jubilados.
- Personas que perciben una pensión distinta de la de jubilación y de prejubilación.
- Personas que altruistamente realizan actividades de tipo benéfico o trabajos sociales.
- Incapacitados para trabajar.
- Otra situación: personas que, sin ejercer ninguna actividad económica, reciben ayuda pública o privada y todas aquellas que no estén incluidas en ninguna de las categorías anteriores, por ejemplo los rentistas.

Por lo que se refiere al ámbito de la encuesta cabe destacar que el ámbito poblacional se refiere a la población que reside en viviendas familiares principales exclusivamente. Se realiza en todo el territorio nacional y con periodicidad trimestral, realizándose las encuestas durante las 13 semanas del trimestre.

El marco de la encuesta viene definido por la división administrativa de España, las 17 comunidades autónomas y las 2 ciudades autónomas. A su vez se dividen las comunidades autónomas en 50 provincias. Partiendo de aquí creamos unas nuevas subdivisiones que llamamos secciones censales

y que son territorios geográficos con unos límites perfectamente definidos, actualizados de manera periódica.

Sin embargo, si queremos hablar del diseño muestral de la encuesta en primer lugar tendremos que mencionar el tipo de muestreo y las unidades muestrales. Se utiliza un muestreo bietápico con estratificación en las unidades de primera etapa exceptuando las ciudades autónomas de Ceuta y Melilla.

Las unidades empleadas en primera etapa son fijas en el tiempo y se corresponden con las secciones censales definidas previamente. Tan solo se extraen de la muestra si entran en el proceso de actualización, siendo estas reemplazadas por otras secciones censales elegidas aleatoriamente.

Las unidades de segunda etapa están constituidas por las viviendas familiares principales, se excluyen entonces segundas residencias y otro tipo de viviendas que no sean de residencia habitual. Por lo que respecta a Ceuta y Melilla en cada una de ellas se realiza un único muestreo aleatorio simple.

Para la estratificación de las unidades de primera etapa se sigue un doble criterio. Geográfico por un lado donde las secciones se agrupan por estratos dentro de cada provincia, siguiendo el tamaño del municipio de pertenencia medido en términos de población.

Por otro lado, se emplea un criterio socioeconómico donde las secciones censales son agrupadas en subestratos dentro de cada estrato, en función de las características socioeconómicas de cada una de ellas.

Ahora presentamos los criterios que se han seguido para la creación de los estratos y subestratos de la muestra.

Los tamaños teóricos para los estratos son los siguientes:

-Estrato 0: Municipios de 500.000 habitantes o más

-Estrato 1: Municipios capital de provincia con menos de 500.000 habitantes.-Estrato 2: Municipios con más de 100.000 habitantes, excepto los anteriores.

-Estrato 3: Municipios de 50.000 a 100.000 habitantes, excepto los anteriores.

-Estrato 4: Municipios de 20.000 a 50.000 habitantes, excepto los

anteriores.

-Estrato 5: Municipios de 10.000 a 20.000 habitantes.

-Estrato 6: Municipios con menos de 10.000 habitantes.

En determinadas provincias se han unificado algunos estratos ya que tenían escasa o nula representación.

Para la creación de los subestratos, dentro de cada estrato se consideran dos grupos de secciones:

-Secciones del estrato 6: perteneciente a municipios pequeños que presentan una variabilidad no demasiado grande en relación con las variables objetivo.

-Resto de secciones: estas secciones se agrupan dentro de sus respectivos estratos mediante el uso de análisis de conglomerados.

Por lo que se refiere al tamaño muestral viene dado por un modelo de optimización matemática, el cual debe cumplir con unos criterios de precisión impuestos por la Unión Europea, estos requisitos de precisión vienen a darnos tres restricciones relacionadas con los siguientes aspectos:

-Requisito de precisión para el paro a nivel nacional.

-Requisito de precisión para el empleo a nivel nacional.

-Requisito de precisión para el paro en cada Comunidad Autónoma.

Para el cálculo del tamaño muestral se lleva a cabo un proceso para la minimización del coste de la encuesta, esto viene a ser la función objetivo de nuestro modelo. En el cual también tenemos unas restricciones para acotar la varianza de los estimadores y evitar que esta se dispare.

La función objetivo viene a ser la siguiente:

$$Q_d(n_d, m^*) = C1n^*(\alpha + \beta m^*) + C2m^* n_d^* f2$$

Donde:

- n_d con $d= 1, \dots, 17$ es el número de secciones censales en la muestra de la comunidad autónoma d , excluyendo Ceuta y Melilla.

- m^* es el número de viviendas seleccionadas de cada sección censal, que se fija como una constante independiente de la comunidad autónoma.
- C_1 = Coste diario medio por investigar una unidad primaria (sección) por CAPI.
- La recta $\alpha + \beta m^*$ = número de días necesarios para investigar m^* viviendas en una sección. Los parámetros α y β son definidos en base a la experiencia adquirida en la recogida en periodos previos de la encuesta.
- C_2 = Coste medio por investigar una vivienda por entrevista CATI.
- f_2 = proporción de viviendas que son investigadas por CATI.

Sumando en todas las comunidades obtenemos la expresión final de la función objetivo:

$$\sum_{d=1}^{19} Q_d (n_d^*, m^*)$$

Los resultados obtenidos de aplicar este modelo de programación matemática arrojan como muestra un total de 5928 secciones censales de las cuales en segunda etapa se seleccionan 13 viviendas por cada una de las secciones. Por lo que respecta a las ciudades autónomas se tiene un muestreo aleatorio simple para cada una de ellas con un tamaño de 260 viviendas principales. Se puede consultar el número de secciones censales por provincia en la tabla presentada en el apéndice de este trabajo.

Además, la afijación adoptada viene dada por la afijación proporcional entre estratos dada por el tamaño poblacional de cada uno de ellos. En cambio, dentro de los propios estratos la afijación entre subestratos es directamente proporcional a la población.

La muestra ha sido seleccionada de tal manera que dentro que cada estrato cualquier vivienda tiene la misma probabilidad de ser seleccionada que las otras, es decir, tenemos muestras autoponderadas dentro de cada estrato. Ya dentro del estrato seleccionado en primera etapa tenemos un número fijo de viviendas a seleccionadas con igual probabilidad mediante la aplicación de un muestreo sistemático con arranque aleatorio.

En estas muestras cada periodo de la encuesta se corresponde con un trimestre, siendo cada sección visitada en una de las 13 semanas del trimestre. La distribución de la muestra es uniforme por lo que en cada provincia el número de secciones por semana es constante.

Posteriormente se procede a la recogida de la información que a día de hoy se realiza del siguiente modo:

- En primeras entrevistas mediante visita personal asistida por ordenador, es decir mediante lo que se denomina método CAPI (Computer Assisted Personal Interviewing).
- En segundas y sucesivas entrevistas (de 2ª a 6ª) mediante llamada telefónica asistida por ordenador, es decir mediante método CATI (Computer Assisted Telephone Interviewing).
- Desde 2020 se permite cumplimentar la encuesta a través de Internet o CAWI (Computer Assisted Web Interviewing)

Las primeras entrevistas se realizan semanalmente por entrevistadores de las 52 delegaciones provinciales del INE. También en las delegaciones provinciales se recoge información de aquellas viviendas que, aun estando en sucesivas entrevistas (de 2ª a 6ª), muestren su preferencia por continuar colaborando mediante visita personal. Las posteriores entrevistas se realizan telefónicamente desde siete centros CATI ubicados a lo largo y ancho del país, repartiéndose entre estos centros la muestra de todo el territorio nacional.

Posteriormente, la información recogida se transmite a Servicios Centrales con sede en Madrid para su tratamiento posterior.

Existen numerosas incidencias que pueden ocurrir a la hora de realizar la encuesta, las más destacables podrían ser el hecho de que la vivienda este vacía o sea de temporada, que los integrantes de esa vivienda estén ausentes o directamente la negativa de estos mismos a la realización de la encuesta.

Por lo contrario, si todo funciona correctamente y se procede a la realización de la encuesta, debemos diferenciar entre sujetos encuestables y no encuestables. Ya sea por la edad, porque no son residentes en dicho domicilio o por cualquier otro motivo contemplado.

Esta información ya recogida ha de ser procesada y tratada, este proceso comprende todos los procesos de depuración e imputación que se realizan con los datos, desde la entrevista hasta la obtención de un fichero final coherente y directamente utilizable para realizar estimaciones.

De cara al tratamiento y proceso de la información debemos tener en cuenta la existencia de las siguientes fases:

Primero pasa por la Delegación Provincial o centro CATI que se encarga de lo siguiente:

- Realización de la entrevista: Desde el cuarto trimestre de 1997, todas las entrevistas de la EPA se realizan mediante ordenadores, lo que ha permitido la detección de errores durante la cumplimentación de dichas entrevistas. Además, esta forma de recogida ha aumentado la calidad de los datos y ha conseguido reducir los errores que se producían con el anterior sistema que se basaba en cuestionarios de papel y grabación centralizada.

- Revisión de la entrevista: Se lleva a cabo un proceso de revisión y cumplimentación de información que haya podido quedar pendiente durante la entrevista.

En segunda instancia el trabajo realizado por los Servicios Centrales viene a ser el siguiente:

- Control de variables de identificación: Una vez completas y revisadas las entrevistas en las delegaciones provinciales o centros CATI, se envían a Servicios Centrales. En una primera fase se revisan, y se depuran siempre que sea necesario, las variables de identificación del hogar y ciertas variables demográficas (como por ejemplo la edad) que serán esenciales para la correcta asignación de los factores de elevación.

- Control de compatibilidad de códigos del resto de variables: A continuación, se eliminan las posibles inconsistencias que pudieran existir entre el resto de variables, de modo que los datos de cada persona queden coherentes.

- Generación de variables derivadas y cálculo de los factores de elevación: Una vez depurados todos los datos, se generan variables derivadas que facilitan la tabulación de los datos y a continuación se calculan los factores de elevación, que permitirán dar estimaciones de la población a partir de

los datos muestrales obtenidos. Además esta variable será de gran interés puesto que la emplearemos en próximos capítulos para los estimadores.

Finalmente, en cuanto a los turnos de rotación de la muestra cabe destacar que de forma indefinida las secciones censales quedan fijas dentro de la muestra, pero las viviendas principales si son renovadas de forma parcial cada trimestre para evitar el cansancio de las familias encuestadas. La renovación se lleva a cabo en una sexta parte de las secciones por lo que la muestra queda dividida en 6 turnos de rotación.

Con todo esto ya queda expuesto los aspectos más relevantes que debemos conocer sobre la EPA, ahora tan solo nos queda conocer brevemente como se evalúa la calidad de los datos que se obtienen y la difusión que esta encuesta tiene por parte del INE.

Antes de conocer como es la evaluación debemos tener claros dos aspectos fundamentales. El primero de ellos es que una estimación será precisa y de calidad si los datos en los que basa también lo son.

En segundo lugar, también es importante saber que el tamaño muestral en el cual están basadas las estimaciones también juega un papel fundamental para la calidad de las mismas. Por esto mismo podemos decir que la calidad de un estimador es una función del error total de la encuesta.

Este error total lo podemos dividir en dos errores, los de muestreo y los ajenos al muestreo.

Por un lado, los errores de muestreo de las estimaciones son calculados de forma trimestral y para la obtención de estos errores se utiliza el método de Jackknife.

Por otro lado, los errores ajenos al muestreo se pueden presentar en cualquier etapa de la encuesta y pueden provenir de alguna de las siguientes fuentes:

- Errores de especificación y medida.

- Errores de marco.

- Errores por falta de respuesta de la unidad, de una o varias preguntas y de respuesta incompleta.

Evaluar estos errores es generalmente muy difícil de llevar a la práctica a la vez que costoso.

Por último, como bien hemos dicho con anterioridad la EPA se trata de una encuesta trimestral por lo que la difusión de los resultados también lo es. Se publican todos los datos de la EPA cada tres meses y se también se proporcionan resultados anualmente a cerca de las variables de submuestra, deciles de salarios del empleo principal y resultados en media aritmética anual de los cuatro trimestres del año.



3-DESCRIPCIÓN DEL FICHERO

En esta sección vamos a ver un breve análisis descriptivo sobre los datos de la EPA del cuarto trimestre de 2021.

Concretamente vamos a conocer los aspectos más relevantes sobre nuestras variables de interés y sobre las que usaremos para crear los dominios.

De inicio contamos con 158290 observaciones de un total de 93 variables, pese a ello tan solo unas pocas de estas las consideraremos de interés.

Por un lado, tenemos la variable sexo que puede ser masculino o femenino, mientras que también tenemos la variable provincia que puede tomar un valor de entre las 50 provincias o las 2 ciudades autónomas que conforman el territorio nacional. Con estas dos variables vamos a construir lo que denominaremos áreas, que no vienen a ser más que las distintas combinaciones de las dos variables, es decir, tendremos un área por cada provincia cruzada con cada sexo. Esto hace un total de 104 áreas.

Ahora bien, antes de poder realizar unas estimaciones debemos conocer bien estas variables y algunas otras para nuestro año de referencia.

Empezamos conociendo la edad de los encuestados en la EPA separando las edades en 5 posibles franjas de edad.

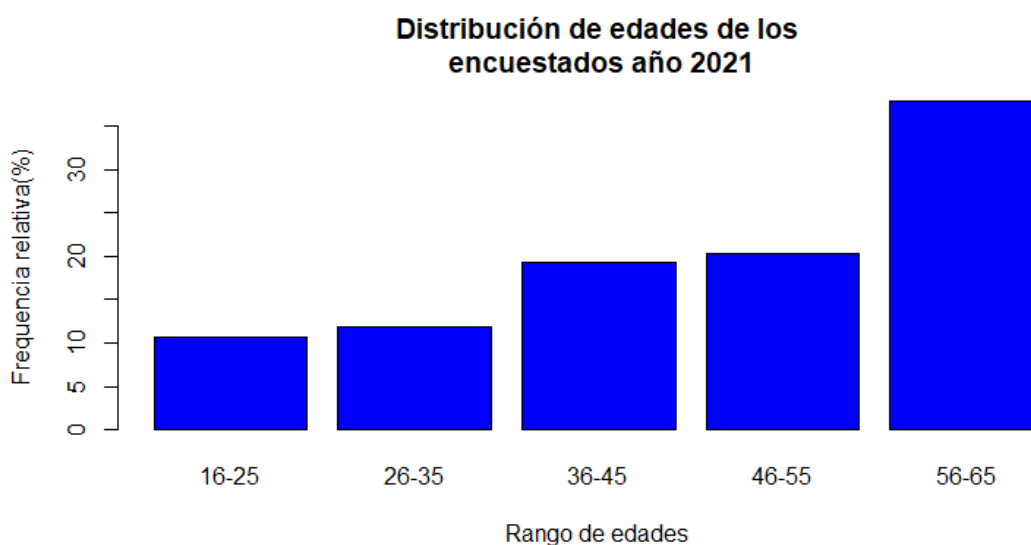


Figura 3.1. Distribución de edades en la muestra del trimestre 1 de 2021.

Se observa como la frecuencia relativa crece a medida que crece la edad del encuestado, es decir, las personas más mayores tienen más presencia en nuestra muestra que las más jóvenes. Se observa en nuestro gráfico de barras como el crecimiento es prácticamente lineal.

La franja que más encuestados tiene es la que va de los 56 a los 65 años mientras que la menos extensa es la franja que va desde los 16 a los 25 años.

Por lo que se refiere al género de los encuestados podemos ver la distribución de estos en el siguiente gráfico circular:

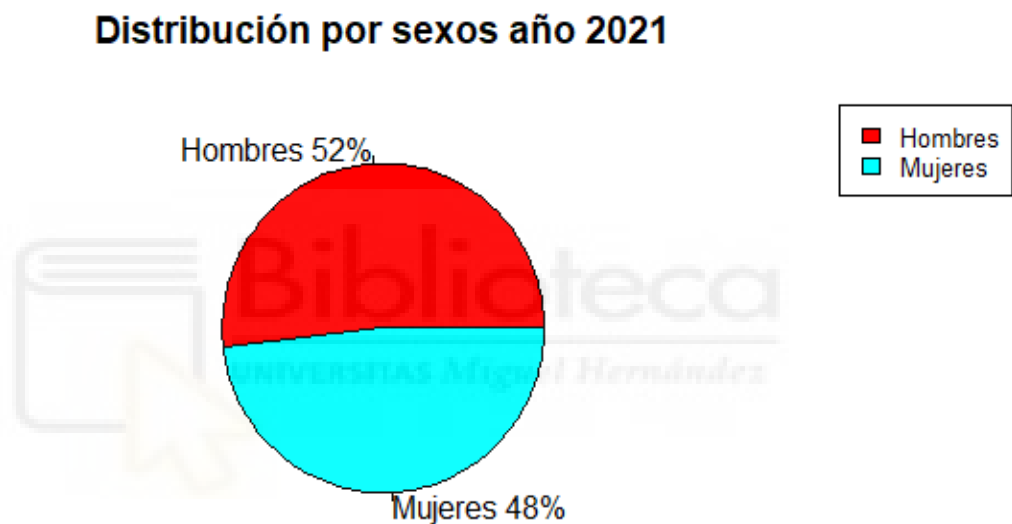


Figura 3.2. Distribución por sexos en la muestra del trimestre 1 de 2021.

Se observa como el número de hombres que participan en la encuesta es ligeramente superior al de mujeres en el periodo de referencia.

Otra variable muy relevante para nuestro estudio y que además será nuestra variable respuesta es la tipología laboral del encuestado. Para crear esta variable hemos tenido que recodificar la variable original (AOI), esta variable recogía 7 opciones laborales que nosotros hemos reducido a tan solo 3. También hemos suprimido los registros que no tenían valor en esta variable. Después de este proceso nos quedan 122057 registros válidos que son los que emplearemos para el cálculo de las estimaciones propuestas con anterioridad.

Ahora podemos ver como ha sido la recodificación que hemos llevado a cabo, que valores hemos suprimido y que nuevos niveles se han creado de nuestra variable:

-Ocupados:

-Ocupados subempleados por insuficiencia de horas

-Resto de ocupados

-Desocupados:

-Parados que buscan primer empleo

-Parados que han trabajado antes

-Inactivos:

-Inactivos 1 (desanimados)

-Inactivos 2 (con los desanimados forman los activos potenciales)

-Inactivos 3 (resto de inactivos)

En el apéndice se puede consultar un extracto del cuestionario utilizado por el INE para la recogida de información sobre esta cuestión y su posterior recodificación.

La Tabla 3.1 recoge el número de ocupados, parados e inactivos, diferenciándolos por sexo en términos absolutos. Los datos se refieren a la muestra EPA del trimestre 1 de 2021.

	OCUPADOS	PARADOS	INACTIVOS	TOTALES
HOMBRES	29065	4416	24659	58140
MUJERES	25754	5247	32916	63917
TOTALES	54819	9663	57575	122057

Tabla 3.1. Frecuencias absolutas de ocupados, parados e inactivos.

La Tabla 3.2 presenta las correspondientes frecuencias relativas de ocupados, parados e inactivos, diferenciándolos por sexo en términos relativos.

	OCUPADOS	PARADOS	INACTIVOS	TOTALES
HOMBRES	23,81	3,62	20,20	47,63
MUJERES	21,10	4,30	26,97	52,37
TOTALES	44,91	7,92	47,17	100,00

Tabla 3.2. Frecuencias relativas de ocupados, parados e inactivos.

Vemos muchas cosas que nos resultan llamativas. Por ejemplo, en el gráfico circular se observa como el número de hombres encuestados es ligeramente superior. Sin embargo, al eliminar los registros sin respuesta vemos como las mujeres pasan a ser mayoría. Esto puede indicar que los hombres tienen mayor facilidad para la no respuesta.

También observamos como el número de ocupados es bastante superior al de parados tanto en hombres como en mujeres. Otra cosa que nos resulta poderosamente llamativa es el número de personas inactivas que en el caso de las mujeres llega a ser ligeramente superior al de ocupadas, mientras que con los hombres es al revés el número de ocupados es también ligeramente superior.

De seguido, las Tabla 3.3 y 3.4 presentan estos mismos resultados, pero tan solo para parados y ocupados que serán los que usemos para el cálculo de nuestras estimaciones.

	OCUPADOS	PARADOS	TOTALES
HOMBRES	29065	4416	33481
MUJERES	25754	5247	31001
TOTALES	54819	9663	64482

Tabla 3.3. Frecuencias absolutas de ocupados y parados.

	OCUPADOS	PARADOS	TOTALES
HOMBRES	45,07	6,85	51,92
MUJERES	39,94	8,14	48,08
TOTALES	85,01	14,99	100,00

Tabla 3.4. Frecuencias relativas de ocupados y parados.

Aquí vemos como la tasa de paro en la muestra está entorno al 15%, con estos datos usando estimadores de dominio estimaremos las tasas de paro diferenciando por sexo y provincia.

Finalmente, otra variable que vamos a emplear y que es muy importante para el cálculo de las estimaciones es la que representa al factor de elevación, esto no viene a ser más que el número de personas de la población a las que representa.

Los factores de elevación se obtienen a partir de las inversas de las probabilidades de inclusión, después de efectuar la corrección por no respuestas (reponderación) y la calibración a totales de provincia-sexo. De ese modo, la suma de los factores de elevación en provincia-sexo coincide con sus totales conocidos.

4-ESTIMADORES DE DOMINIO

En este capítulo exponemos brevemente la razón de emplear estimadores de dominio y presentamos la formulación de los estimadores empleados para la realización de nuestro estudio.

En esta ocasión contamos con una muestra de la población española que viene a ser la EPA, esta muestra la utilizaremos para estimar parámetros en una variedad de subpoblaciones que es a lo que nosotros denominaremos dominios. Estos usualmente se definen como grupos socioeconómicos o áreas en nuestro caso las provincias cruzadas con el sexo.

Si estos dominios son pequeños, es decir, no son lo suficientemente grandes como para obtener estimadores directos que sean lo suficientemente precisos se les suele denominar áreas pequeñas. En cambio si se trata del caso contrario decimos que el dominio es grande.

Estos estimadores son directos cuando están basados exclusivamente en los datos específicos del dominio. Se trata de un estimador basado en el diseño muestral y por lo tanto emplean los pesos muestrales. Por último, estos estimadores también se pueden hacer valer de información auxiliar.

Ahora que ya conocemos los estimadores de dominio, estamos en disposición de presentar los estimadores que emplearemos y su formulación matemática.

La notación empleada es la siguiente:

-Índices: s se usa para la muestra, $d=1, \dots, D$ para las áreas pequeñas (provincias cruzadas con sexo), $j=1; \dots, N$ para los individuos.

-Población y muestra: $P = U_{d=1}^D P_d$ para la población y $s = U_{d=1}^D s_d$ para la muestra.

-Tamaños: N para la población y n para la muestra.

-Probabilidades de inclusión: $\pi_j = P(\text{seleccionar la unidad } j), j=1, \dots, n.$

- Pesos teóricos del diseño muestral: $1/\pi_j, j=1, \dots, n.$

- Factores de elevación: $w_j, j=1, \dots, n$. Los factores de elevación se obtienen corrigiendo los pesos teóricos por la falta de respuesta y posteriormente calibrándolos a cantidades de provincia-sexo conocidas.

En nuestro caso las variables son las siguientes:

$$y_j = \begin{cases} 1 & \text{si el individuo } j \text{ es parado,} \\ 0 & \text{en caso contrario.} \end{cases}$$

$$z_j = \begin{cases} 1 & \text{si el individuo } j \text{ es ocupado,} \\ 0 & \text{en caso contrario.} \end{cases}$$

Pretendemos estimar el número total de parados, número total de ocupados y la tasa de paro por áreas. Con la notación introducida, estos indicadores socio económicos son

$$Y_d = \sum_{j \in P_d} y_j \quad Z_d = \sum_{j \in P_d} z_j$$

$$R_d = \frac{Y_d}{Y_d + Z_d} = \frac{\bar{Y}_d}{\bar{Y}_d + \bar{Z}_d}$$

donde

$$\bar{Y}_d = Y_d / N_d,$$

$$\bar{Z}_d = Z_d / N_d, \text{ y}$$

N_d es el número de individuos en el área d .

El estimador directo del total Y_d es:

$$\hat{Y}_d^{\text{direct}} = \sum_{j \in S_d} w_j y_j$$

La varianza del estimador directo del total es la siguiente:

$$V_{\pi}[\hat{Y}_d^{direct}] = \sum_{i=1}^{N_d} \sum_{j=1}^{N_d} (\pi_{ij} - \pi_i \pi_j) \frac{y_j y_j}{\pi_j \pi_j}$$

En condiciones ideales de inexistencia de no respuestas, se puede usar el estimador

$$\hat{V}_{\pi}[\hat{Y}_d^{direct}] = \sum_{i \in S_d} \sum_{j \in S_d} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij}} \right) \frac{y_j y_j}{\pi_j \pi_j}$$

Sin embargo, en la práctica se usa un estimador que no requiere usar las probabilidades de inclusión doble y emplea solamente los factores de elevación. El estimador es

$$\hat{V}_{\pi}[\hat{Y}_d^{direct}] = \sum_{i \in S_d} w_j (w_j - 1) y_j^2.$$

5-ESTIMADORES APLICADOS A LA EPA (T4 21)

En este apartado pasamos a presentar los estimadores obtenidos al implementar la formulación matemática anteriormente presentada.

Para ello utilizaremos la representación gráfica tanto de las tasas de paro como de los coeficientes de variación mediante mapas, todos ellos para el conjunto de provincias diferenciadas por sexo.

En el apéndice podemos encontrar las tablas donde se detallan todos los resultados obtenidos para cada una de las provincias, se puede consultar el número de parados, número de ocupados, tasa de paro y coeficiente de variación de la tasa de paro.

También la tabla que representa las medias de las tasas de paro por comunidad, así como el número total estimado de parados y ocupados para cada sexo y cada comunidad autónoma.

Hay que tener en cuenta que la media de la tasa de paro por comunidad no es equivalente con la tasa de paro de dicha comunidad. La tasa de paro por comunidad autónoma es la media ponderada de la tasa de paro de sus provincias, mientras que aquí presentamos otro indicador que no tiene en cuenta el tamaño de las provincias y que puede ser útil para ser comparado con la tasa de paro de toda la comunidad, de modo que podamos ver si las grandes provincias tienden a estar más o menos desempleadas que las pequeñas.

Otro factor a tener en cuenta es que los datos de los totales de ocupados y parados son ocupados por lo que están altamente influenciados por el tamaño de la región de pertenencia.

En primera instancia vamos a presentar las tasas de paro, en esta ocasión para las mujeres:

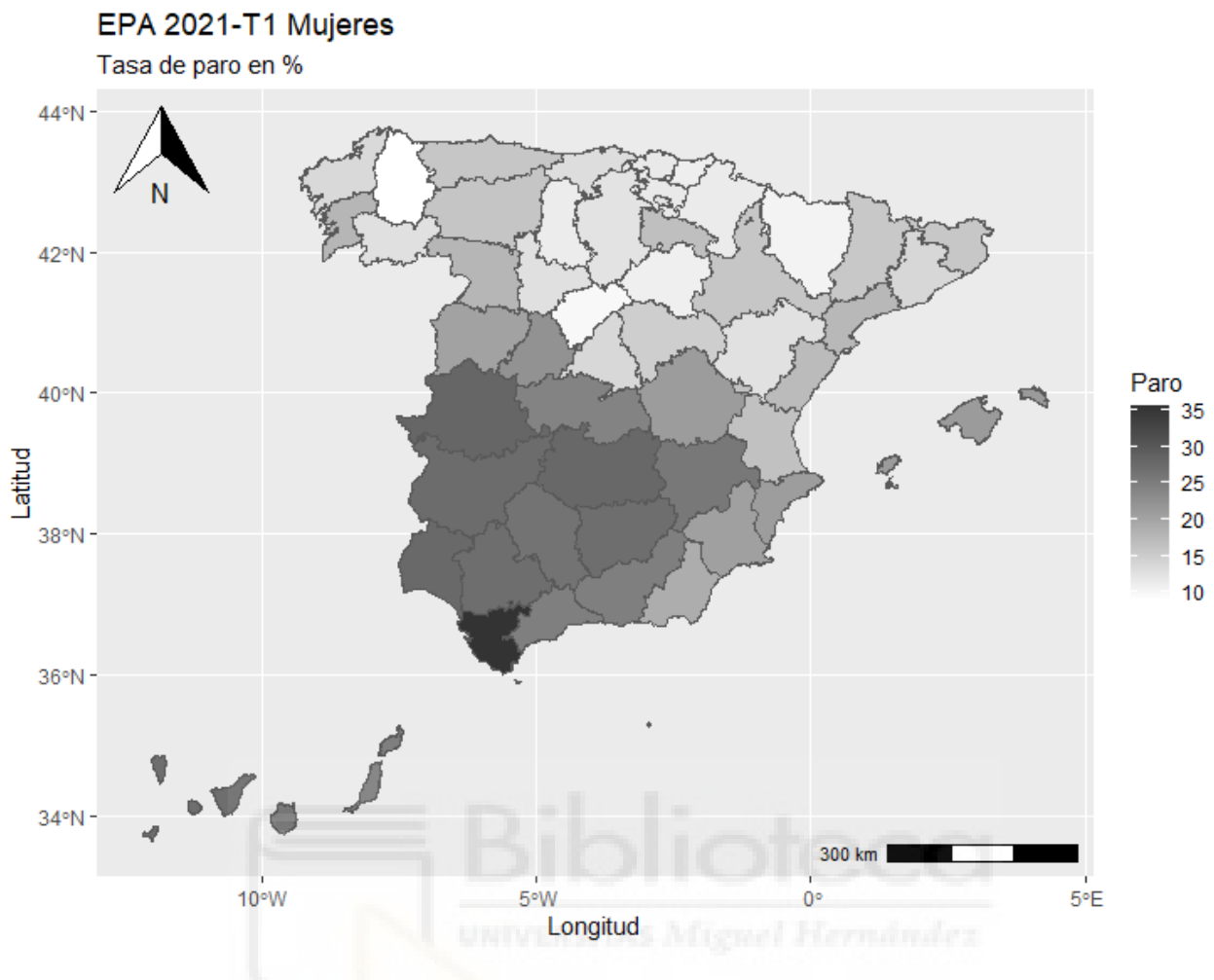


Figura 5.1. Representación de la tasa de paro femenina en porcentaje a nivel nacional con datos del primer trimestre de 2021.

Las comunidades más castigadas por el paro femenino son Andalucía, Extremadura y las ciudades autónomas, siendo Cádiz con un 35,43% de tasa paro la más alta de todo el territorio nacional y siendo la única junto con Ceuta en superar el 30% de paro.

En el otro extremo encontramos a Lugo como la provincia con menos tasa de paro femenino 8,98% seguido de cerca por Segovia, Navarra y las tres comunidades pertenecientes al País Vasco, siendo las dos primeras las únicas en bajar del 10%.

La media de las tasas de paro de todas las 52 regiones incluidas en el estudio es del 18,76% y presentando una variabilidad bastante grande entre las distintas regiones que componen España.

A continuación, presentamos el mapa de tasas de paro para hombres perteneciente al primer periodo del año 2021:

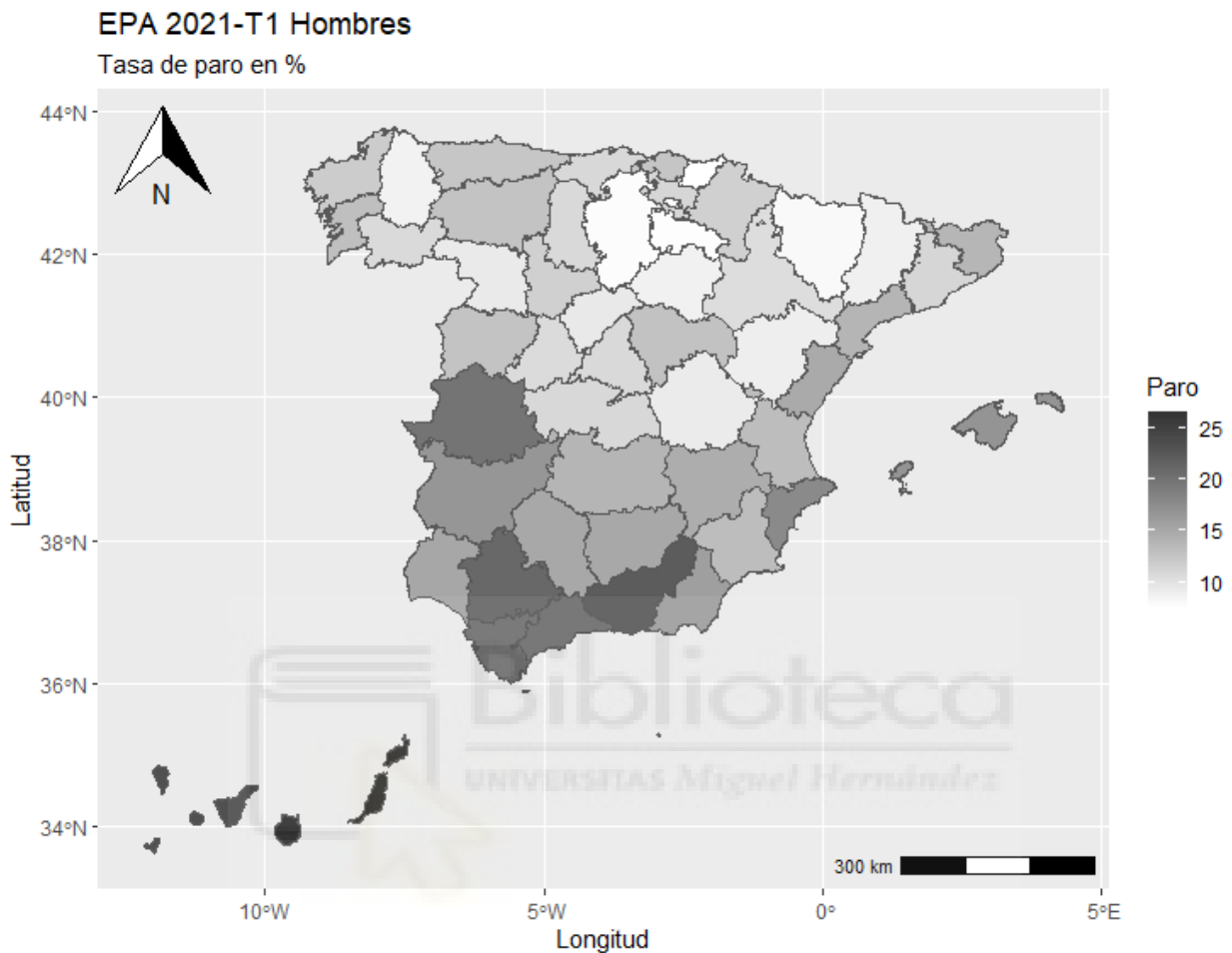


Figura 5.2. Representación de la tasa de paro masculina en porcentaje a nivel nacional con datos del primer trimestre de 2021.

Vemos marcadas similitudes entre ambos mapas sobre todo geográficamente hablando, las regiones con altas tasas de paro para mujeres también las suelen presentar en el caso de los hombres.

También es cierto que claramente los porcentajes de paro para hombres son inferiores a los de las mujeres de hecho, existe una diferencia de casi 10 puntos entre el máximo de unas estimaciones y de otras.

La mayor tasa de paro la presenta Las Palmas de Gran Canaria con un 26,5% seguida de cerca por Santa Cruz de Tenerife, Ceuta, Granada y Sevilla. Se

observa como las islas Canarias en su conjunto son las que presentan peores datos.

Como nota positiva podemos destacar las provincias de Guipúzcoa (7,63%), La Rioja (7,8%) y Burgos (7,9%) entre otras, siendo estas las únicas tres provincias con porcentajes inferiores a los 8 puntos.

Por último, en cuanto a las medias de las tasas de paro la comunidad que mejores datos presenta es La Rioja. La media del conjunto de España para hombres es del 13,69% por lo que hay una diferencia de unos 5 puntos en entre la tasa de paro en hombres y mujeres.

Ahora que ya conocemos lo más relevante en cuanto a las tasas de paro, vamos a completar el conocimiento de estas viendo los coeficientes de variación para cada una de las áreas.

Para ello, vamos a presentar dos mapas uno por cada sexo donde podamos ver los coeficientes de variación para cada área y así poder encontrar algún factor relevante en el estudio.

Empezamos presentando los coeficientes de variación de las estimaciones hechas para los 52 territorios incluidos en el estudio y que en este caso se corresponden con las estimaciones pertenecientes al primer trimestre del 2021 para las mujeres.

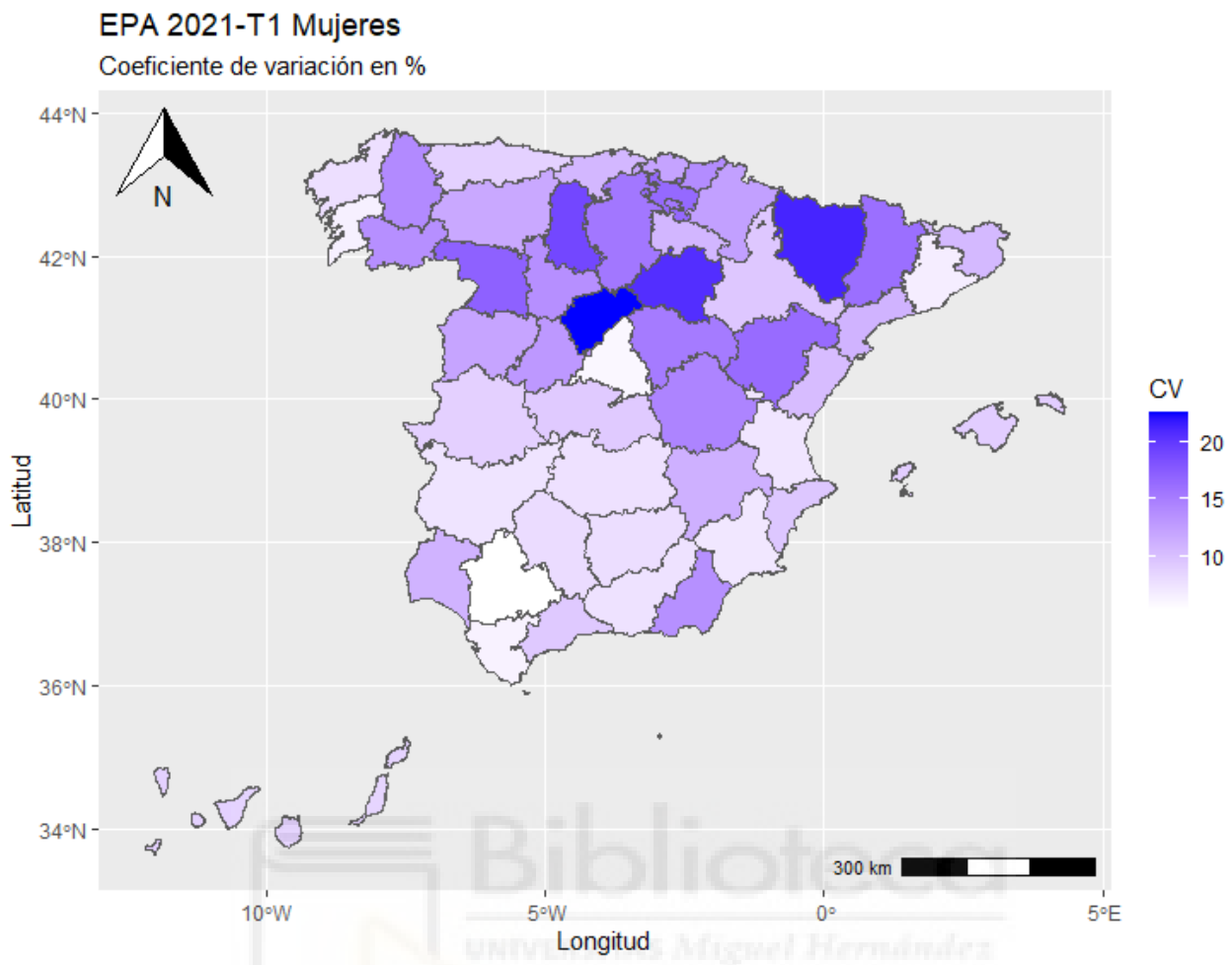


Figura 5.3. Representación del coeficiente de variación para la tasa de paro femenina en porcentaje a nivel nacional con datos del primer trimestre de 2021.

La variación media ha sido del 11,61% estando los valores comprendidos entre los 5,37 y 22,47 puntos porcentuales.

Los valores más altos que son los que vemos más oscuros en el mapa vienen a ser Segovia, Soria y Huesca que son las únicas provincias además que presentan valores por encima del 20%.

Sin embargo, los valores más pequeños que aparecen en blanco en el mapa se corresponden con Sevilla y Madrid. También presentan valores bajos en este coeficiente provincias como Valencia, Barcelona o Pontevedra.

El último mapa que presentamos seguido se corresponde con los coeficientes de variación para hombres

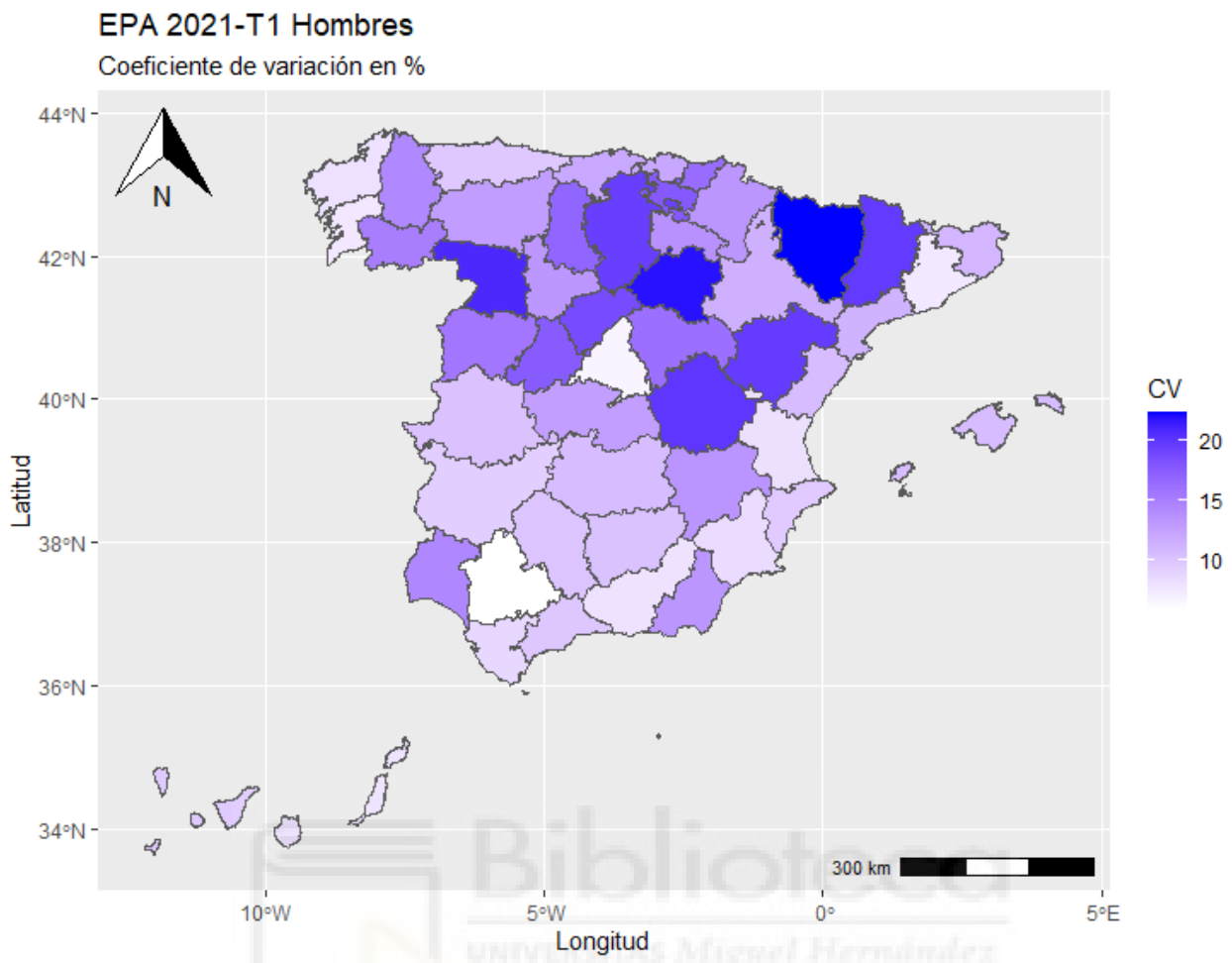


Figura 5.4. Representación del coeficiente de variación para la tasa de paro femenina en porcentaje a nivel nacional con datos del primer trimestre de 2021.

Al igual que con las tasas de paro vemos como los resultados para hombres y mujeres son geográficamente muy similares, pero a diferencia de las tasas de paro no existe una diferencia tan sustancial en el valor de los coeficientes de variación para ambos sexos.

Si bien es cierto que en este caso la media de los coeficientes de variación es del 13,07% ligeramente superior al del sexo femenino que tenía su media en el 11,67%.

El valor más alto lo presenta Huesca (22,22%) mientras que el valor más pequeño y por lo tanto menos variable lo presenta también la provincia de Sevilla (5,88%). Es reseñable también que los territorios que anteriormente

presentaban valores muy pequeños en este coeficiente de variación siguen teniendo valores bajos, cosa que indica poca variación en los datos de estas provincias.



6-CONCLUSIONES

Aquí se presentan las conclusiones derivadas del capítulo anterior. En primer lugar presentaremos las conclusiones generales y luego las conclusiones extraídas para cada caso.

A nivel general podemos decir que no hay diferencias geográficamente relevantes entre sexos, ni para la tasa de paro ni para el coeficiente de variación. Esto viene a decir que una provincia si presenta un alto nivel de paro para los hombres generalmente también lo va a presentar para las mujeres y viceversa.

Si hablamos en términos relativos a tasas de paro se observa un claro patrón geográfico que podríamos denominarlo factor meridional, ya que las zonas del norte del país presentan tasas de paro sustancialmente más pequeñas que las demás sobre todo que las zonas del sur que es donde se encuentran las tasas de paro más altas.

En este sentido, pero en menor medida también se observa como las regiones más occidentales presentan una mayor tasa de paro que las ubicadas en el este peninsular, por lo que las zonas ubicadas en el suroeste tienden a ser las que presentan mayores tasas de paro.

También cabe destacar que las mujeres presentan tasas de paro más altas que los hombres y se observa perfectamente en el mapa este hecho que hemos denominado factor meridional.

Por otro lado, en lo que se refiere a los coeficientes de variación vemos como ambos mapas son muy parecidos, aunque para el sexo masculino se tienen ligeramente mayores coeficientes de variación. Las zonas que presentan colores más oscuros en un mapa también los presentan en el otro, esto indica que las zonas que presentan mucha variabilidad en los datos en un sexo también lo presentan para el otro y viceversa.

Finalmente, otro factor a destacar es el de la naturaleza de las provincias que presentan mayores coeficientes de variación y por tanto mayor variabilidad. Al observar a las mismas vemos como las provincias más pequeñas son las que presentan mayor variabilidad, esto es lógico ya que

al ser las más pequeñas tienen un menor número de secciones censales asignadas y por lo tanto menor número de datos.

En el caso contrario de provincias grandes con mayor número de secciones censales vemos como los coeficientes de variación son mucho más pequeños, puesto que tienen un mayor número de datos y no son tan susceptibles a la presencia de valores extremos.

Para concluir podríamos valorar generalmente las tasas de paro como altas, la existencia de tasas de paro por encima del 20% e incluso del 30% es un factor preocupante para la economía de un país independientemente del motivo de estas cifras.

Si bien es cierto que no todas las provincias presentan cifras altas, sí que a nivel global y en comparación con las tasas de otros países de la Unión Europea las tasas son bastante altas.

Obviamente otro aspecto llamativo es la variabilidad existente entre áreas ya que no existe una uniformidad en cuanto a tasas de paro se refiere, de hecho existen provincias que duplican e incluso triplican los porcentajes de paro de otras comunidades.

Por destacar un último aspecto podríamos decir que un factor explicativo de estas tasas de paro tanto a nivel de cifras como a nivel de zonas, podría ser el período de los datos. España presenta una estructura laboral muy estacional y los datos del primer trimestre del año, pueden generar esas cifras un poco más altas en las zonas más dependientes del turismo.

7-IMPLEMENTACIÓN DE LOS ESTIMADORES EN RSTUDIO

Como hemos comentado con anterioridad el software empleado para la implementación de los estimadores y la obtención de los resultados se ha trabajado con el lenguaje de programación R bajo el entorno de Rstudio.

En este capítulo procedemos a mostrar cómo ha sido este proceso mostrando directamente la programación de los estimadores, así como los pasos intermedios necesarios. Para ello mostraremos el código empleado más relevante con extractos de este y se describirá que es lo que se logra con cada función empleada.

El archivo que hemos creado para la realización del trabajo se denomina “EPA.R”, por la extensión se sobreentiende que se trata de nuestro archivo base en lenguaje de R donde vamos a realizar todos los pasos para la obtención de los estimadores.

Junto con este archivo también disponemos del archivo de datos descargado del portal de datos abiertos del INE, se trata de un archivo en formato CSV denominado “EPA_2021T1.csv” donde se encuentran todas las respuestas después de haber realizado todas las encuestas de las encuestas de las que se compone la EPA.

Por último, también empleamos un archivo de geodatos obtenido de un portal abierto de datos que aparece mencionado en la bibliografía. Este archivo contiene la información de todo el territorio español (comunidades autónomas, provincias y ciudades autónomas) con sus respectivas coordenadas, lo que nos permite poder graficar el mapa para posteriormente pintarlo con las estimaciones obtenidas.

A continuación, pasamos a presentar detalladamente los pasos más importantes que se han llevado a cabo para el trabajo.

El primero paso es la lectura del fichero de la EPA que contiene todos los datos.

```

7 setwd("C:/Users/paper/Desktop/TFG/datos_1t21/CSV/")
8 dat<-read.csv('EPA_2021T1.csv', header = TRUE, sep = "")
9 head(dat, 10)
10
11 datcol <- ncol(dat); datcol           # número de columnas (variables) en dat
12 n <- nrow(dat); n                     # número de filas (casos) in dat
13 narea <- length(unique(dat$PROV)); narea # número de areas
14 nsex <- length(unique(dat$SEXO1)); nsex  # número de categorías en sexo

```

Figura 7.1. Lectura del fichero 'EPA_2021T1.csv'.

Como se puede observar primero establecemos el directorio de trabajo donde tenemos almacenados todos los ficheros de datos. Luego leemos el fichero que anteriormente hemos seleccionado, comprobamos algunos aspectos para ver que se ha leído correctamente y sacamos información relevante del conjunto de datos.

```

> datcol <- ncol(dat); datcol           # número de columnas (variables) en dat
[1] 91
> n <- nrow(dat); n                     # número de filas (casos) in dat
[1] 142850
> narea <- length(unique(dat$PROV)); narea # número de areas
[1] 52
> nsex <- length(unique(dat$SEXO1)); nsex  # número de categorías en sexo
[1] 2

```

Figura 7.2. Comprobación de la correcta lectura de los datos.

En esta captura podemos ver como los datos han sido leídos correctamente. Además, sabemos que contamos con un conjunto de datos de 91 variables, 142850 registros y como es lógico disponemos de dos sexos y 52 provincias.

A continuación, vamos a mostrar cómo sería la vista de los datos una vez que ya han sido leídos y almacenados en Rstudio. Se muestra tan solo un fragmento de las 12 primeras variables y los 18 primeros registros del total.

	CICLO	CCAA	PROV	NVIVI	NIVEL	NPERS	EDAD1	RELPP1	SEXO1	NCONY	NPADRE	NMADRE
1	194	16	1	1	1	1	40	1	Hombres	2	0	0
2	194	16	1	1	1	2	30	2	Mujeres	1	0	0
3	194	16	1	2	1	1	35	1	Mujeres	2	0	0
4	194	16	1	2	1	2	30	2	Hombres	1	0	0
5	194	16	1	2	2	3	0	3	Mujeres	0	2	1
6	194	16	1	3	1	1	65	1	Hombres	2	0	0
7	194	16	1	3	1	2	60	2	Mujeres	1	0	0
8	194	16	1	3	1	3	25	3	Mujeres	0	1	2
9	194	16	1	4	1	1	45	1	Mujeres	2	0	0
10	194	16	1	4	1	2	45	2	Hombres	1	0	0
11	194	16	1	4	1	3	16	3	Hombres	0	2	1
12	194	16	1	5	1	1	50	2	Hombres	2	0	0
13	194	16	1	5	1	2	50	1	Mujeres	1	0	0
14	194	16	1	5	2	3	10	3	Mujeres	0	1	2
15	194	16	1	6	1	1	40	1	Mujeres	0	0	0
16	194	16	1	7	1	1	45	1	Hombres	2	0	0
17	194	16	1	7	1	2	35	2	Mujeres	1	0	0
18	194	16	1	7	2	3	10	3	Mujeres	0	1	2

Figura 7.3. Vista de los datos una vez leídos en Rstudio.

El siguiente paso sería retocar los datos que queremos cambiar, tratar con los datos en blanco y crear nuevas variables que vayan a ser necesarias.

```

46 # empleados, parados e inactivos
47
48 library(dplyr)
49 library(tidyverse)
50
51 datos=dat %>%
52   mutate(AOI=fct_collapse(factor(AOI),Ocupados=c("4","3"),Parados=c("6","5"),
53     Inactivos=c("7","8","9")),SEXO1=fct_collapse(factor(SEXO1),Hombres=c("1"),Mujeres=c("6")))
54
55 table(datos$SEXO1,datos$AOI)

```

Figura 7.4. Recodificación de la variable AOI.

En la anterior imagen podemos ver como hemos recodificado las variables AOI y SEXO. Con AOI hemos convertido los valores 3 y 4 en ocupados, los valores 5 y 6 en parados y por último los valores 7, 8 y 9 en inactivos. Mientras que para SEXO hemos recodificado el valor 1 en hombres y el valor 6 en mujeres.

Posteriormente creamos la tabla 3.1. que nos indica el número total de parados, ocupados e inactivos.

El siguiente paso sería crear el conjunto de datos que vamos a emplear con tan solo 7 variables, esto lo hacemos por comodidad pues es más sencillo trabajar con 7 variables que con 91. Ahora bien, necesitamos variables binarias que identifiquen el sexo del encuestado y su estado laboral para ello usamos las variables indicadoras (dummies).

Finalmente con nuestro conjunto de datos definitivo eliminamos los valores en blanco o alternativamente en R NA's (del inglés, non available).

```
63 dat<-dummy_cols(datos, ignore_na=TRUE, remove_first_dummy=FALSE, remove_selected_columns = TRUE)
64 z<-dat[,2]
65 a<-dat[,3]
66 b<-dat[,88]
67 c<-dat[,104:106]
68 w<-dat[,87]
69 dat<-data.frame(z,a,b,c,w)
70 names(dat)<-c("Comunidad","Provincia","Sexo","Ocupados","Parados","Inactivos","w")
71 dat<-na.omit(dat)
```

Figura 7.5. Creación del archivo de datos definitivo con las variables de interés y eliminación de NA's.

Ya disponemos de los datos y el formato de estos correcto para poder realizar las estimaciones, pero antes tenemos que renombrar algunas variables que emplearemos para el cálculo de estos estimadores (se puede consultar su formulación en el cuarto capítulo).

```
74 # Renombramos algunas variables
75
76 y1 <- dat$Parados; # AOI Desempleados
77 y2 <- dat$Ocupados; # AOI Ocupados
78 y3<- dat$Inactivos; # Inactivos
79 area <- dat$Provincia; # PROV
80 sex <- dat$Sexo # SEX01
81 ccaa<- dat$Comunidad
82 w<- dat$w
```

Figura 7.6. Renombramiento de variables.

Ahora que ya tenemos todos los pasos previos necesarios podemos proseguir con el cálculo de las estimaciones.

En primer lugar, vamos a estimar los totales de parados y totales de ocupados. Luego creamos objeto donde vamos a guardar estas dos estimaciones junto con la comunidad autónoma, la provincia y el sexo de cada una de ellas.

Para ello utilizamos la función *aggregate* que va a ser muy recurrente durante todo el trabajo, esta función nos permite aplicar funciones sobre los subgrupos creados según las variables de agrupación en nuestro caso sexo y provincia por las cuales se quieren hacer las agrupaciones

```
83 # estimación de totales de parados y ocupados
84 dir1.ds <- aggregate(w*data.frame(y1,y2), by=list(Area=ccaa,Area=area,Sex=sex), sum)
85 # Assign column names
86 names(dir1.ds) <- c("ccaa","area", "sex", "y1tot", "y2tot")
```

Figura 7.7. Estimación de totales de parados y ocupados.

El siguiente paso sería estimar tanto las varianzas como los coeficientes de variación de parados y ocupados para más tarde añadir estas estimaciones a nuestro objeto creado anteriormente donde vamos a recoger todos los resultados que vayamos obteniendo.

```
90 ## Estimación de varianzas y coeficientes de variación
91
92 vardir1.ds <- aggregate(w*(w-1)*data.frame(y1^2,y2^2),
93                       by=list(Area=ccaa,Area=area,Sex=sex), sum)
94 # Assign column names
95 names(vardir1.ds) <- c("ccaa","area", "sex", "y1var", "y2var")
96 # Add columns y1var and y2var
97 dir1.ds <- cbind(dir1.ds, vardir1.ds$y1var, vardir1.ds$y2var)
98 # CV for y1
99 y1cv <- 100*sqrt(vardir1.ds$y1var)/abs(dir1.ds$y1tot)
100 # CV for y2
101 y2cv <- 100*sqrt(vardir1.ds$y2var)/abs(dir1.ds$y2tot)
102 # Add columns y1cv and y2cv
103 dir1.ds <- cbind(dir1.ds, y1cv, y2cv)
104 # Change column names for dir1.ds
105 namesds <- c("ccaa","area", "sex", "y1tot", "y2tot", "y1var",
106            "y2var", "y1cv","y2cv")
107 names(dir1.ds) <- namesds
```

Figura 7.8. Estimación de varianzas y coeficientes de variación.

Cuando ya tengamos los totales y las varianzas podemos estimar las tasas de paro que va a ser nuestra principal variable de respuesta. Para ello como se puede ver en el capítulo cuarto, simplemente dividimos los parados entre el sumatorio de estos y los ocupados y luego multiplicamos por 100.

```
110 ## Estimación de tasas de paro
111
112 dirrate.ds <- 100*dir1.ds$y1tot/(dir1.ds$y1tot + dir1.ds$y2tot)
113 dir1.ds <- cbind(dir1.ds, rate=dirrate.ds)
```

Figura 7.9. Estimación de tasas de paro.

Ahora que ya tenemos la tasa de paro que viene ser lo más relevante de este estudio, procedemos a calcular las estimaciones de su varianza y su coeficiente de variación. Estos resultados son unos de los que hemos mapeado en el quinto capítulo para poder ver de forma rápida los coeficientes de variación en las distintas provincias.

```
124 ## Varianza de la tasa de paro
125
126 s1.ds <- dir1.ds$y2tot^2*dir1.ds$y1var/(dir1.ds$y1tot+dir1.ds$y2tot)^4
127 s2.ds <- dir1.ds$y1tot^2*dir1.ds$y2var/(dir1.ds$y1tot+dir1.ds$y2tot)^4
128 s12.ds <- 2*dir1.ds$y1tot*dir1.ds$y2tot*covardir1.ds$covar/
129 (dir1.ds$y1tot + dir1.ds$y2tot)^4
130
131 # Estimación de varianza y coeficiente de variación
132 dir1.ds$vrate <- 10^4*(s1.ds+s2.ds-s12.ds)
133 dir1.ds$cvrate <- 100*sqrt(dir1.ds$vrate)/abs(dir1.ds$rate)
```

Figura 7.10. Estimación de varianzas y coeficientes de variación para la tasa de paro.

Ahora que ya tenemos todos los estimadores tan solo nos faltaría por conocer el total de parados y ocupados, pero ahora ya si diferenciados por sexo y por provincia. Como es evidente este resultado estará enormemente condicionado por el tamaño de la provincia en cuestión.

```
142 ## Calculo de los totales de parados y ocupados por provincia y sexo
143
144 dir <- aggregate(w*data.frame(1/dat$w,1,y1,y2), by=list(Area=ccaa,Area=area,Sex=sex), sum)
145 # Column names
146 names(dir) <- c("ccaa","area", "sex", "nds", "hatNds", "y1tot", "y2tot")
```

Figura 7.11. Estimación de totales de parados y ocupados por sexo y provincia.

Ya tenemos todos los estimadores y resultado que pretendíamos obtener al principio del trabajo y que planteamos en el cuarto capítulo. Por lo que el siguiente paso es presentar los resultados de una forma clara y eficaz. Con esta finalidad vamos a crear unos mapas donde de un vistazo podamos conocer la mayor información posible.

Para crear estos mapas vamos a emplear las librerías *ggplot*, *ggspatial* y *sf* que nos van a permitir graficar esta información de forma sencilla.

Como siempre el primer paso es leer los datos necesarios, en este caso tenemos que leer el fichero de geodatos donde está guardada la información de toda España y que nos va a permitir dibujarla para posteriormente añadir la capa de datos que queramos, en nuestro caso tasas de paro y coeficientes de variación.

```
171 setwd("C:/Users/paper/Desktop/TFG/")
172 provincias<- st_read("provincias-espanolas.geojson")
173
174 # Ordenando por la columna 'codigo'
175 provincias<- provincias[order(provincias$codigo), ]
```

Figura 7.12. Lectura del archivo de geodatos.

En este archivo las provincias no están ordenadas por su código identificativo mientras que nuestros resultados si lo están. Esto es un serio problema ya que cuando hacemos el gráfico con *ggplot* esta función empareja objetos del mismo tamaño por el orden en el que están dentro del mismo, es decir, si no ordenamos el objeto provincias por su código en orden ascendente el gráfico no será correcto puesto que se pintarán tasas de paro en provincias equivocadas. Para esto utilizamos la función *order* que nos va a permitir que el emparejamiento entre la tasa de paro y la provincia a la que hace referencia esa tasa sea el adecuado.

El siguiente problema que se nos presenta no es ni más ni menos que la propia geografía en sí, esto viene a ser que cuando realizamos el gráfico la distancia entre las islas Canarias y el resto de la península es excesivo y no permite que se visualice bien la información.

Para solventar esto tenemos que acercar de alguna forma las islas al resto de territorios y esto lo conseguimos modificando las coordenadas de Tenerife y de Las Palmas de la siguiente manera:

```
182 #acercamos canarias
183 provincias$geometry[[35]]<-provincias$geometry[[35]]+6
184 provincias$geometry[[38]]<-provincias$geometry[[38]]+6
```

Figura 7.13. Cambio de coordenadas de Canarias para aproximarlas al resto de territorios.

Finalmente, ya disponemos de las estimaciones y de los geodatos en un formato correcto por lo que solo nos queda el último el paso que es juntar estos dos archivos para conseguir uno de nuestros mapas.

A continuación, se presenta tan solo uno de los cuatro mapas realizados concretamente se trata de la figura 5.2. ya que todos ellos presentan la misma estructura, tan solo cambian los datos que pretendemos graficar.

En este caso se trata de la tasa de paro masculina:

```
198 Paro<-output1[53:104,8]
199
200 mapahombres<- ggplot(data = provincias) +
201   geom_sf(aes(fill = Paro))+
202   xlab("Longitud") + ylab("Latitud") +
203   ggtitle("EPA 2021-T1 Hombres",
204           subtitle = "Tasa de paro en %")+
205   annotation_scale(location='br') +
206   annotation_north_arrow(location='tl')
207
208 mapahombres + scale_fill_gradient(low='white', high='grey20')
```

Figura 7.14. Gráfico de la tasa de paro masculina perteneciente al primer periodo de 2021.

Podemos ver como primero seleccionamos la capa de datos de la tasa de paro para hombres que la tenemos almacenada en el objeto *output1*, concretamente es la octava variable y se corresponde con los registros que van del 53 al 104.

Posteriormente, creamos el mapa con la función *ggplot* y añadimos los geodatos, más tarde agregamos al objeto principal que es el anteriormente descrito la capa de datos con la que queremos rellenar el mapa, en este caso la tasa de paro masculina y por último especificamos y agregamos algunos aspectos estéticos como pueden ser los colores, título y subtítulo, flecha de norte y escala entre otros.



8-APÉNDICE A: TABLAS

En este apéndice presentamos algunas tablas referidas durante el estudio y que aquí aparecen detalladas.

Pretendemos que este apéndice sea un lugar de consulta donde poder conocer las estimaciones obtenidas con exactitud

En primer lugar, podemos ver la tabla citada en el segundo capítulo que representa el número de secciones censales por cada comunidad autónoma:

COMUNIDAD AUTÓNOMA	SECCIONES
Andalucía	751
Aragón	230
Principado de Asturias	188
Islas Baleares	191
Islas Canarias	239
Cantabria	159
Castilla y León	531
Castilla – La Mancha	336
Cataluña	490
Comunidad Valenciana	366
Extremadura	210
Galicia	576
Comunidad de Madrid	325
Región de Murcia	168
Com. Foral de Navarra	174
País Vasco	239
La Rioja	125
Ceuta (*)	260
Melilla (*)	260

* viviendas principales

Tabla A.1. Número de secciones censales pertenecientes a cada Comunidad Autónoma.

La siguiente tabla muestra la distribución de estas secciones censales por provincia y por tipo de estrato al que pertenecen, estos estratos aparecen definidos en el segundo capítulo.

ESTRATO								
	0	1	2	3	4	5	6	Total
01 Araba/Álava		37				4	8	49
02 Albacete		27			13		24	64
03 Alicante/Alacant		24	15	29	37	11	18	134
04 Almería		20		18		17	18	73
05 Ávila		16					29	45
06 Badajoz		27		11	17	10	59	124
07 Balears, Illes		66		17	53	29	26	191
08 Barcelona	89		59	43	43	25	31	290
09 Burgos		33			13		22	68
10 Cáceres		19			15		52	86
11 Cádiz		10	28	33	17	8	9	105
12 Castellón/Castelló		20		5	19	5	17	66
13 Ciudad Real		12			22	17	26	77
14 Córdoba		31			16	9	21	77
15 Coruña, A		48	18	14	44	26	60	210
16 Cuenca		12					33	45
17 Girona		9			23	13	26	71
18 Granada		23			17	18	28	86
19 Guadalajara		16				12	22	50
20 Gipuzkoa		20		7	15	21	16	79
21 Huelva		17			17	9	18	61
22 Huesca		12				13	23	48
23 Jaén		12		5	13	13	24	67
24 León		23			23		37	83
25 Lleida		17				8	31	56
26 Rioja, La		59			9	14	43	125
27 Lugo		28	0	0	0	16	48	92
28 Madrid	170		71	41	15	10	18	325
29 Málaga	44		11	38	19		21	133
30 Murcia		51	24	18	47	28		168
31 Navarra		55			20	26	73	174
32 Orense		30	0	0	0	14	46	90
33 Asturias		40	50	23	19	28	28	188
34 Palencia		22					23	45
35 Palmas, Las		42	11	20	32	10	7	122
36 Pontevedra		16	58	0	40	40	30	184
37 Salamanca		29				7	28	64
38 Santa Cruz de Tenerife		24	17	22	30	8	16	117
39 Cantabria		48		14	23	17	57	159

40 Segovia		15					30	45
41 Sevilla	55		10	9	29	23	23	149
42 Soria		19					26	45
43 Tarragona		12	9		21	6	25	73
44 Teruel		12					33	45
45 Toledo		12		19		15	54	100
46 Valencia/València	52			19	46	19	30	166
47 Valladolid		54			10		27	91
48 Bizkaia		35	10	7	26	13	20	111
49 Zamora		16				4	25	45
50 Zaragoza	97					11	29	137
Total (*)	507	1170	391	412	803	577	1438	5298

Tabla A.2. Número de secciones censales por provincia y por cada tipo de estrato.

A continuación, presentamos las dos tablas que contienen todos los valores estimados para cada una de las áreas, tanto para totales de parados, totales de ocupados, tasas de paro y coeficiente de variación.

Presentamos una tabla para el sexo femenino y otra para el masculino, estos resultados están presentados y detallados en el quinto capítulo.

COMUNIDAD	PROVINCIA	PARADOS	OCUPADOS	% PARO	%CV
País Vasco	Álava	8441	64756,38	11,53	16,48
Castilla - La Mancha	Albacete	22569,92	65371,3	25,66	11,20
Comunitat Valenciana	Alicante	83598,29	318526,1	20,79	9,49
Andalucía	Almería	28509,48	122814,22	18,84	13,59
Castilla y León	Ávila	6962,88	24308,88	22,27	12,95
Extremadura	Badajoz	38025,97	101345,56	27,28	7,34
Illes Balears	Islas Baleares	61781,84	231634,64	21,06	9,00
Cataluña	Barcelona	190911,1	1203746,58	13,69	6,70
Castilla y León	Burgos	9511,02	67978,71	12,27	15,46
Extremadura	Cáceres	21553,96	55132,27	28,11	8,77
Andalucía	Cádiz	88665,95	161602,01	35,43	6,33
Comunitat Valenciana	Castellón	20773,14	101062,1	17,05	10,26
Castilla - La Mancha	Ciudad Real	27705,09	72202,02	27,73	7,58
Andalucía	Córdoba	43664,99	122660,92	26,25	7,92
Galicia	La Coruña	33203,31	212939	13,49	7,67
Castilla - La Mancha	Cuenca	7781,87	29437,76	20,91	14,44
Cataluña	Gerona	29575,83	159217,89	15,67	10,51

Andalucía	Granada	47672,63	145144	24,72	7,49
Castilla - La Mancha	Guadalajara	8613,1	49575,76	14,8	15,17
País Vasco	Guipúzcoa	18421,38	149840,11	10,95	13,87
Andalucía	Huelva	28358,51	74022,78	27,7	10,96
Aragón	Huesca	4850,85	41651,03	10,43	21,12
Andalucía	Jaén	31948,51	86510,64	26,97	7,80
Castilla y León	León	13940,16	74109,87	15,83	11,77
Cataluña	Lleida	16416,97	89760,76	15,46	16,04
Rioja, La	La Rioja	12024,03	60587,09	16,56	10,75
Galicia	Lugo	6521,69	66097,6	8,98	14,05
Comunidad de Madrid	Madrid	237111,6	1514214,68	13,54	5,99
Andalucía	Málaga	92308,63	281120	24,72	9,34
Región de Murcia	Murcia	65892,72	255751,38	20,49	7,14
Comunidad Foral de Navarra	Navarra	16698,26	131077,75	11,3	12,40
Galicia	Orense	7871,9	54454,67	12,63	13,60
Asturias, Principado de	Asturias	34360,49	183659,11	15,76	8,73
Castilla y León	Palencia	3711,99	29452,83	11,19	18,94
Canarias	Las Palmas	65692,58	199782,75	24,75	8,55
Galicia	Pontevedra	36052,49	168144,09	17,66	6,49
Castilla y León	Salamanca	14272,36	56532,93	20,16	12,04
Canarias	Santa Cruz de Tenerife	68488,59	185080,72	27,01	8,75
Cantabria	Cantabria	16504,64	110986,17	12,95	10,67
Castilla y León	Segovia	3094,41	28505,71	9,79	22,47
Andalucía	Sevilla	115190,1	310982,49	27,03	5,37
Castilla y León	Soria	2085,41	17082,46	10,88	20,60
Cataluña	Tarragona	31522	149471,64	17,42	10,97
Aragón	Teruel	3523,66	24778,06	12,45	16,29
Castilla - La Mancha	Toledo	36538,1	114185,05	24,24	9,20
Comunitat Valenciana	Valencia	96351,82	491198,08	16,4	7,27
Castilla y León	Valladolid	14566,2	100732	12,63	13,62
País Vasco	Vizcaya	28776,64	221382,96	11,5	12,15
Castilla y León	Zamora	5869,36	27302,49	17,69	17,04
Aragón	Zaragoza	35261,13	188670,41	15,75	9,46
Ceuta	Ceuta	5246,06	11448,41	31,42	15,45
Melilla	Melilla	4718,63	13691,3	25,63	16,47

Tabla A.3. Totales de empleados, totales de parados, tasa de paro y coeficiente de variación para mujeres de todos los territorios durante el primer trimestre de 2021.

Ahora presentamos la tabla A.4. donde podemos consultar la misma información que en la anterior, pero para el sexo masculino.

COMUNIDAD	PROVINCIA	PARADOS	OCUPADOS	% PARO	%CV
País Vasco	Álava	8878,81	71103,2	11,1	17,65
Castilla - La Mancha	Albacete	14897,47	86678,1	14,67	13,43
Comunitat Valenciana	Alicante	85683,6	395170,23	17,82	9,58
Andalucía	Almería	33008,8	174495,78	15,91	13,24
Castilla y León	Ávila	4180,9	35054,13	10,66	17,48
Extremadura	Badajoz	28314,61	141250,59	16,7	9,31
Illes Balears	Islas Baleares	55024,62	270392,47	16,91	10,53
Cataluña	Barcelona	161920,9	1297707,83	11,09	7,46
Castilla y León	Burgos	7415,25	86498,6	7,9	19,39
Extremadura	Cáceres	18712,1	75353,56	19,89	10,12
Andalucía	Cádiz	61959,21	244554,28	20,21	8,68
Comunitat Valenciana	Castellón	21657,86	123957,46	14,87	10,56
Castilla - La Mancha	Ciudad Real	17875,9	110208,85	13,96	10,49
Andalucía	Córdoba	30919,04	174356,94	15,06	9,93
Galicia	La Coruña	31262,01	230373,58	11,95	7,98
Castilla - La Mancha	Cuenca	5031,05	49159,19	9,28	19,93
Cataluña	Gerona	28071,29	174966,21	13,83	10,96
Andalucía	Granada	53729,72	188246,07	22,2	7,92
Castilla - La Mancha	Guadalajara	9667,75	65766,33	12,82	16,07
País Vasco	Guipúzcoa	13313,43	161158,14	7,63	16,31
Andalucía	Huelva	19575,9	109562,03	15,16	14,36
Aragón	Huesca	4919,57	55836,99	8,1	22,22
Andalucía	Jaén	23956,41	135124,77	15,06	10,16
Castilla y León	León	13401,15	92201,11	12,69	12,81
Cataluña	Lleida	10792,75	110543,93	8,89	19,71
Rioja, La	La Rioja	6445,06	76210,96	7,8	13,61
Galicia	Lugo	6731,52	70947,72	8,67	14,31
Comunidad de Madrid	Madrid	192665,8	1593661,73	10,79	6,66
Andalucía	Málaga	84043,08	329271,04	20,33	9,79
Región de Murcia	Murcia	54509,65	357125,75	13,24	8,31
Comunidad Foral de Navarra	Navarra	19190,93	146585,88	11,58	13,25
Galicia	Orense	7115,34	59403,47	10,7	15,00
Asturias, Principado de	Asturias	28651,54	200695,02	12,49	9,73
Castilla y León	Palencia	4050,24	33370,44	10,82	16,81
Canarias	Las Palmas	74646,6	207033,84	26,5	7,84
Galicia	Pontevedra	29556,28	197401,97	13,02	7,52
Castilla y León	Salamanca	9809,46	67630,89	12,67	15,54
Canarias	Santa Cruz de Tenerife	63845,43	208176,43	23,47	9,51
Cantabria	Cantabria	15659,41	127516,55	10,94	11,96
Castilla y León	Segovia	4056,34	38681,23	9,49	18,60
Andalucía	Sevilla	104438,6	391796,48	21,05	5,88
Castilla y León	Soria	2016,32	20960,14	8,78	21,78
Cataluña	Tarragona	30514,47	188445,33	13,94	11,20
Aragón	Teruel	3005,57	30311,56	9,02	19,59

Castilla - La Mancha	Toledo	21016,69	171500,8	10,92	12,62
Comunitat Valenciana	Valencia	86198,42	563847,45	13,26	8,04
Castilla y León	Valladolid	15483,23	118039,76	11,6	13,20
País Vasco	Vizcaya	33421,29	233140,46	12,54	12,08
Castilla y León	Zamora	3782,6	36642,71	9,36	20,75
Aragón	Zaragoza	26406,23	225227,85	10,49	11,42
Ceuta	Ceuta	4932,21	14065,34	25,96	17,18
Melilla	Melilla	3871,45	17632,48	18	21,18
				13,69	13,07

Tabla A.4. Totales de empleados, totales de parados, tasa de paro y coeficiente de variación para hombres de todos los territorios durante el primer trimestre de 2021.

Ahora presentamos la tabla donde podemos ver la media de las tasas de paro por comunidades autónomas, no se trata de una media ponderada que podríamos tomar como tasa de paro de dicha comunidad autónoma

COMUNIDAD AUTÓNOMA	MUJERES	HOMBRES
Andalucía	26,46	18,12
Aragón	12,88	9,20
Principado de Asturias	15,76	12,49
Islas Canarias	25,88	24,99
Cantabria	12,95	10,94
Castilla - La Mancha	22,67	12,33
Castilla y León	14,75	10,44
Cataluña	15,56	11,94
Ceuta	31,42	25,96
Comunidad de Madrid	13,54	10,79
Comunidad Foral de Navarra	11,3	11,58
Comunidad Valenciana	18,08	15,32
Extremadura	27,7	18,30
Galicia	13,19	11,09
Islas Baleares	21,06	16,91
Melilla	25,63	18,00
País Vasco	11,33	10,42
Región de Murcia	20,49	13,24
La Rioja	16,56	7,80
TOTAL GENERAL	18,76	13,69

Tabla A.5. Medias de las tasas de paro de cada provincia agrupadas por comunidad autónoma para los datos del primer trimestre del año 2021.

Para terminar con las tablas incluidas en este apéndice, presentamos dos tablas en las que se pueden observar las estimaciones obtenidas tanto de totales de parados como de totales de ocupados. Se presentan dos tablas, una para cada sexo y en ellas podemos ver estos totales para cada comunidad autónoma.

En primer lugar presentamos la que hace referencia al mercado laboral femenino:

COMUNIDAD AUTÓNOMA	PARADOS	OCUPADOS
Andalucía	476319	1304857
Aragón	43636	255100
Asturias, Principado de	34360	183659
Canarias	134181	384863
Cantabria	16505	110986
Castilla - La Mancha	103208	330772
Castilla y León	74014	426006
Cataluña	268426	1602197
Ceuta	5246	11448
Comunidad de Madrid	237112	1514215
Comunidad Foral de Navarra	16698	131078
Comunitat Valenciana	200723	910786
Extremadura	59580	156478
Galicia	83649	501635
Illes Balears	61782	231635
Melilla	4719	13691
País Vasco	55639	435979
Región de Murcia	65893	255751
Rioja, La	12024	60587
Total general	1953713	8821724

Tabla A.6. Totales de parados y ocupados en el mercado laboral femenino para cada comunidad autónoma durante el primer trimestre de 2021.

Y en segundo lugar la tabla que hace referencia al mercado laboral masculino:

COMUNIDAD AUTÓNOMA	PARADOS	OCUPADOS
Andalucía	411631	1747407
Aragón	34331	311376
Asturias, Principado de	28652	200695
Canarias	138492	415210
Cantabria	15659	127517
Castilla - La Mancha	68489	483313
Castilla y León	64195	529079
Cataluña	231299	1771663
Ceuta	4932	14065
Comunidad de Madrid	192666	1593662
Comunidad Foral de Navarra	19191	146586
Comunitat Valenciana	193540	1082975
Extremadura	47027	216604
Galicia	74665	558127
Illes Balears	55025	270392
Melilla	3871	17632
País Vasco	55614	465402
Región de Murcia	54510	357126
Rioja, La	6445	76211
Total general	1700234	10385044

Tabla A.7. Totales de parados y ocupados en el mercado laboral masculino para cada comunidad autónoma durante el primer trimestre de 2021.

Finalmente, presentamos la parte del cuestionario de la EPA realizado por el INE donde se puede observar cómo se pregunta al encuestado por su situación laboral.

9-APÉNDICE B: CUESTIONARIO EPA

B. RELACIÓN CON LA ACTIVIDAD ECONÓMICA

Si EDAD >89 o EDAD=15 pasar a F

Las preguntas siguientes tratan sobre su situación laboral

1. La semana de referencia, ¿trabajó aunque sólo fuera una hora?
- Sí 1
 - No 6 Pasar a AYUDFA
 - No sabe 0 Pasar a AYUDFA

TRA

2. ¿Se le paga por ese trabajo u obtiene un beneficio económico con él?
- Sí 1 Pasar a C
 - No 6
 - No sabe 0

REM

3. Si TRA=6,0:
La semana de referencia, ¿ayudó en la empresa, negocio o explotación de un familiar sin que le paguen por ello?
Si REM=6,0:
¿Fue un trabajo no remunerado en la empresa, negocio o explotación de un familiar (ayuda familiar)?
- Sí 1
 - No 6
 - No sabe 0

AYUDFA

4. Si TRA=1:
La semana de referencia ¿tenía un empleo o negocio, aunque no trabajara en él esa semana?
(Se refiere a empleos remunerados y a negocios en los que se trabaje habitualmente al menos 1 hora a la semana. Incluidos ayudas familiares)
Si TRA=6:
A pesar de no haber trabajado la semana de referencia ¿tenía un empleo o negocio?
(Se refiere a empleos remunerados y a negocios en los que se trabaje habitualmente al menos 1 hora a la semana. Incluidos ayudas familiares)
Si TRA=0:
La semana de referencia ¿tenía un empleo o negocio?
(Se refiere a empleos remunerados y a negocios en los que se trabaje habitualmente al menos 1 hora a la semana. Incluidos ayudas familiares)
- Sí 1
 - No 6
 - No sabe 0

AUSENT

Si AYUDFA=1, pasar a C
Si AUSENT=6,0, pasar a D

5. ¿Cuál es la razón principal por la que no trabajó en dicho empleo?
(No se leen las opciones, se señala la que indique el entrevistado. Si dice varias, se le pide la principal. Sólo si dice que no sabe, se leen las opciones. Si aún así no sabe, se marca esa opción)
- Vacaciones o días de permiso 1 Pasar a SITU11
 - Permiso por nacimiento de un hijo 2 Pasar a SITU11
 - Excedencia por cuidado de hijos 3 Pasar a SITU11
 - Enfermedad, accidente o incapacidad temporal del encuestado 4 Pasar a SITU11
 - Jornada de verano, horario variable, flexible o similar 5 Pasar a SITU11
 - Actividades de representación sindical 6 Pasar a SITU11
 - Nuevo empleo en el que aún no había empezado a trabajar 7 Pasar a D
 - Fijo discontinuo o trabajador estacional en la época de menor actividad 8
 - Mal tiempo 9 Pasar a SITU11
 - Paro parcial por razones técnicas o económicas 10 Pasar a SITU11
 - Se encuentra en expediente de regulación de empleo 11 Pasar a SITU11
 - Huelga o conflicto laboral 12 Pasar a SITU11
 - Haber recibido enseñanza o formación relacionada con el trabajo 13 Pasar a SITU11
 - Razones personales o responsabilidades familiares 14 Pasar a SITU11
 - Otras razones 15 Pasar a SITU11
 - No sabe 0 Pasar a SITU11

RZNOTB

6. Aun estando en la época de menor actividad, ¿realiza regularmente alguna tarea relacionada con ese empleo o negocio estacional?
- Sí 1
 - No 6 Pasar a D
 - No sabe 0 Pasar a D

ESTAC

7. **En su trabajo principal, ¿cuál de las siguientes es su situación profesional?** SITU11
 Le leo las opciones:
- Miembro de una cooperativa 1
 - Empresario con asalariados 2 Pasar a cuadro 1 antes de REMEMP
 - Trabajador independiente o empresario sin asalariados 3 Pasar a DESAC1
 - Ayuda en la empresa o negocio familiar 4 Pasar a cuadro 1 antes de REMEMP
 - Asalariado sector público (es decir, de algún tipo de administración o ente público) 5 Pasar a cuadro 1 antes de REMEMP
 - Asalariado sector privado 6 Pasar a cuadro 1 antes de REMEMP
 - Otra situación (especificar: _____) 7 Pasar a cuadro 1 antes de REMEMP
 - No sabe 0 Pasar a cuadro 1 antes de REMEMP

8. **¿Es además asalariado en ella?** COOAS1
 - Sí 1 Pasar a cuadro 1 antes de REMEMP
 - No 6 Pasar a cuadro 1 antes de REMEMP
 No sabe 0 Pasar a cuadro 1 antes de REMEMP

9. **¿Cómo desarrolla su actividad?** Le leo las opciones:
(Entrevistador: lea las opciones hasta que dé una respuesta afirmativa)
 - Por su propia cuenta en su negocio, buscando o atendiendo a sus clientes 1 Pasar a cuadro 1 antes de REMEMP
 - Tiene un cliente casi en exclusiva aunque también atiende otros clientes cuando surgen 2 Pasar a cuadro 1 antes de REMEMP
 - En o para la empresa o negocio de un empleador que le proporciona trabajo 3 Pasar a cuadro 1 antes de REMEMP
 No sabe 0 Pasar a cuadro 1 antes de REMEMP

Cuadro 1:

Si RZNOTB=1,2,4,5,13 o ESTAC=1 pasar a C

Cuadro 2:

Si RZNOTB<=>3 pasar a REINC

10. Si RZNOTB=3 y SITU11=4 REMEMP
¿Recibe algún tipo de remuneración relacionada con esta excedencia?
 Si RZNOTB=3 y SITU11=5,6,7,0
¿Sigue recibiendo alguna remuneración?
 Si RZNOTB=3 y SITU11=1-3
¿Continúa percibiendo beneficios, ganancias o remuneración de su negocio o empresa?
 - Sí 1 Pasar a C
 - No 6
 No sabe 0

11. **¿Cuándo se va a reincorporar a su trabajo, contando desde el inicio de la ausencia?** REINC
 Le leo las opciones:
 Si RZNOTB=11
 En el caso de que haya salido y entrado en el ERTE varias veces, tenga en cuenta que la duración se refiere al período desde la última entrada
- En un período inferior o igual a 3 meses 1 Pasar a C **
 - En un período superior a 3 meses 2
 - No tiene seguridad de volverse a reincorporar 3
 - No sabe 0

Pasar a D

**Para mantener el criterio homogéneo durante todo el período de pandemia COVID 19, los asalariados ausentes de su empleo por estar en expediente de regulación, con una ausencia total de más de 3 meses y cobrando más del 50% del salario se tratan, a efectos de cumplimentación del cuestionario, como si el período total de la ausencia fuera inferior o igual a 3 meses, lo que equivale a considerar que en esas condiciones su vínculo con el empleo es fuerte.

10-REFERENCIAS

La bibliografía consultada para la realización del trabajo ha sido la siguiente:

-MORALES GONZÁLEZ, Domingo. Muestreo en poblaciones finitas. Septiembre 2020.

-INE [en línea].[consulta: 22 julio 2021]. Disponible en: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176918&menu=ultiDatos&idp=1254735976595

-OPENDATASOFT [en línea].[consulta: 22 julio 2021]. Disponible en: https://public.opendatasoft.com/explore/conjunto_de_datos/provincias-espanolas/table/?sort=provincia

-SMALL AREA ESTIMATION [en línea].[consulta: 17 junio 2021]. Disponible en: <https://github.com/small-area-estimation/sae-book>

-MAPPINGGIS [en línea].[consulta: 17 junio 2021]. Disponible en: <https://mappinggis.com/2019/07/creacion-de-mapas-con-r-y-ggplot2/>