

# Estadísticas de población de Andalucía basadas en registros administrativos

## Ficheros de microdatos: estructura y diccionario de datos

### Tabla de contenido

1. Características generales.....	2
2. Diseño de registro del fichero muestral.....	3
3. Variables y características empleadas.....	3
3.1. Claves numéricas de identificación de hogares y personas.....	3
3.2. Variables geográficas utilizadas.....	4
3.3. Variables personales básicas.....	5
3.4. Variables temáticas referidas a la persona.....	6
3.4.1. Información de pensiones contributivas.....	6
3.4.2. Información de afiliación a la Seguridad Social.....	7
3.4.3. Información de demandantes de empleo.....	8
3.4.4. Información de titulares de MUFACE.....	10
4. Control del riesgo de revelación y preservación del secreto estadístico.....	10

## 1. Características generales

A partir del marco poblacional integrado anual generado en las Estadísticas de población de Andalucía basadas en registros administrativos (**EPABRA**), y tras su depuración estadística para suprimir inconsistencias, se ha realizado un producto de difusión consistente en un **fichero muestral de microdatos integrados**, con fecha de referencia 1 de enero de 2019, preparados para suministrar a los usuarios la información necesaria que puedan requerir en sus respectivos proyectos de análisis e investigación.

Del fichero marco poblacional integrado anual se ha extraído una muestra de viviendas familiares (no se han considerado elegibles las viviendas calificadas como establecimientos colectivos). Cada vivienda seleccionada aporta a la muestra todas las personas que son miembros de la misma. El procedimiento de selección de las viviendas se ha realizado mediante muestreo estratificado utilizando como estratos las zonas geográficas resultantes de delimitar en cada provincia las unidades territoriales formadas por los municipios pertenecientes a cada uno de los seis niveles de tamaño municipal contemplados: capital de provincia; municipios con más de 100.000 habitantes (no capital de provincia); municipios de 50.001 a 100.000 habitantes; municipios de 20.001 a 50.000 habitantes; municipios de 10.001 a 20.000 habitantes; y municipios de hasta 10.000 habitantes. En cada uno de los estratos se ha realizado un muestreo aleatorio simple de las unidades muestrales, con una fracción de muestreo del 10%, aproximadamente, en cada estrato.

Además del fichero muestral de microdatos integrados con fecha de referencia 1 de enero de 2019 se han elaborado otros tres ficheros muestrales adicionales. Estos ficheros se corresponden con la información de los tres trimestres anteriores, con fechas de referencia 1 de abril de 2018, 1 de julio de 2018 y 1 de octubre de 2018.

La extracción de estas tres muestras trimestrales de microdatos se lleva a cabo tomando como marco poblacional el último fichero muestral de microdatos integrados producto de la operación anual, es decir, el de fecha de referencia 1 de enero de 2019. Esto implica que, a diferencia del procedimiento del muestreo anual antes descrito, no se emplea como unidad de muestreo la vivienda familiar, sino que para las muestras trimestrales se seleccionan aquellos individuos presentes en la muestra de 1 de enero de 2019 que se encuentren a su vez recogidos en los marcos poblacionales integrados trimestrales correspondientes a las fechas de referencia 1 de abril de 2018, 1 de julio de 2018 y 1 de octubre de 2018.

Este procedimiento permite practicar un seguimiento retrospectivo a la población de la muestra anual de 1 de enero de 2019 en los tres trimestres previos, manteniéndose fija la información recogida a tal fecha de referencia en las variables Clave identificadora de persona (B02), Año de nacimiento (B03), Sexo (B04), Nivel educativo (B09) y Clave de vivienda familiar (B10), e informando de los valores que toman las restantes variables de la muestra en las fechas de referencia trimestrales de 1 de abril de 2018, 1 de julio de 2018 y 1 de octubre de 2018.

El seguimiento retrospectivo en esta serie de conjuntos de microdatos simplifica el estudio de la evolución a lo largo del año de las variables de marcado carácter estacional, como las relativas a la relación de la población con la actividad económica. De tal modo, al conocer las transiciones que se producen entre los diferentes estados de dichas variables, se posibilita el

estudio de las características de los colectivos de individuos que cambian, por ejemplo, de estar parados a estar afiliados o de estar afiliados a estar jubilados.

Desde un punto de vista analítico, los ficheros muestrales de microdatos integrados tienen por objetivo proporcionar una base de datos de la que derivar distribuciones relativas de las distintas combinaciones de los valores de las variables disponibles.

## 2. Diseño de registro del fichero muestral

Nombre	Inicio	Ancho	Contenido
<b>B. Información básica</b>			
B01	1	8	Fecha de referencia de la información
B02	9	7	Clave numérica identificadora de persona
B03	16	4	Año de nacimiento
B04	20	1	Sexo
B05	21	3	Nacionalidad
B06	24	2	Provincia de residencia
B07	26	1	Tamaño del municipio de residencia
B08	27	1	Grado de urbanización de la celda de 1 km de residencia
B09	28	2	Nivel educativo
B10	30	7	Clave numérica identificadora de vivienda
B11	37	1	Relación entre el lugar de nacimiento y el de residencia
B12	38	4	Año de llegada al municipio de residencia
<b>P. Información de pensiones contributivas</b>			
P01	42	1	Número de prestaciones percibidas
P02	43	1	Tipología de la prestación
P03	44	1	Importe mensual bruto acumulado
<b>A. Información de afiliación a la Seguridad Social</b>			
A01	45	1	Número de afiliaciones
A02	46	3	Régimen de afiliación a la Seguridad Social
A03	49	1	Actividad económica del régimen o de la cuenta de cotización
A04	50	4	Tipo de contrato
<b>D. Información de demandantes de empleo</b>			
D01	54	1	Clasificación de la demanda
D02	55	2	Ocupación preferente en la que el demandante desea trabajar
D03	57	1	Tiempo que lleva inscrita la demanda
D04	58	1	Ámbito territorial de búsqueda de empleo
D05	59	1	Actividad económica en la que ha trabajado anteriormente
<b>M. Información de titulares de MUFACE</b>			
M01	60	1	Situación administrativa
M02	61	1	Actividad económica en la que trabaja (o ha trabajado)

## 3. Variables y características empleadas

### 3.1. Claves numéricas de identificación de hogares y personas

En el fichero muestral de microdatos se incluyen las variables B02 y B10 que contienen las claves identificativas de persona y vivienda, respectivamente. Al tratarse de una muestra de

viviendas familiares, la variable B10 permite hacer una agrupación de todas las personas que habitan en la misma vivienda a fecha de referencia 1 de enero de 2019.

B01. Fecha de referencia de la información contenida en el fichero en formato AAAAMMDD.

B02. Clave numérica identificadora de persona.

B10. Clave numérica identificadora de vivienda.

### 3.2. Variables geográficas utilizadas

B06. Provincia de residencia.

Posibles valores:

04	Almería
11	Cádiz
14	Córdoba
18	Granada
21	Huelva
23	Jaén
29	Málaga
41	Sevilla

B07. Tamaño del municipio de residencia.

Posibles valores:

1	Capital de provincia
2	Más de 100.000 habitantes (no capital de provincia)
3	De 50.001 a 100.000 habitantes
4	De 20.001 a 50.000 habitantes
5	De 10.001 a 20.000 habitantes
6	Hasta 10.000 habitantes

B08. Grado de urbanización de la celda de 1 km de residencia.

En esta variable se definen distintas categorías de áreas basándose en criterios de contigüidad geográfica, densidad y umbrales de población según la metodología descrita por EUROSTAT para la asignación de las tipologías territoriales en las celdas de la malla estadística de 1 km<sup>2</sup>. La fecha de referencia de la clasificación utilizada es 1 de enero de 2018.

Los valores tomados por la variable son:

Valor	Descripción	Significado
1	Centro urbano	Conglomerado de celdas contiguas de 1 km <sup>2</sup> con una densidad de al menos 1.500 habitantes por km <sup>2</sup> y que conjuntamente agrupan un mínimo de población de 50.000 habitantes
2	Agrupación urbana	Conglomerado de celdas contiguas de 1 km <sup>2</sup> con una densidad de al menos 300 habitantes por km <sup>2</sup> y que conjuntamente agrupan un mínimo de población de 5.000 habitantes

3	Celda en malla rural	El resto de celdas no clasificadas en alguno de los apartados anteriores
0	Sin celda asignada	

---

### 3.3. Variables personales básicas

B03. Año de nacimiento

B04. Sexo

Posibles valores:

- 1 Hombre
- 6 Mujer

B05. Nacionalidad

Posibles valores:

- 100 España
- 211 Unión Europea (excluida España)
- 212 Resto de Europa
- 220 África
- 231 América del norte
- 232 Centro América y Caribe
- 233 Sudamérica
- 240 Asia
- 250 Oceanía
- 260 Apátridas

B09. Nivel educativo

Posibles valores:

- 00 No consta/No procede
- 01 No sabe leer ni escribir
- 02 Sin estudios o enseñanza primaria incompleta
- 03 Bachiller elemental, EGB completa, graduado escolar, ESO o equivalente
- 04 Formación profesional de primer grado, oficialía industrial
- 05 Formación profesional de segundo grado, maestría industrial
- 06 Bachiller superior, BUP
- 07 Títulos de estudios superiores no universitarios
- 08 Diplomatura en escuelas universitarias, arquitectura o ingeniería técnica
- 09 Licenciatura universitaria, arquitectura o ingeniería superior
- 10 Doctorado y estudios de postgrado o especialización para licenciados

B11. Relación entre el lugar de nacimiento y el de residencia

Posibles valores:

- 1 Mismo municipio
- 2 Otro municipio de la misma provincia
- 3 Otro municipio de otra provincia andaluza

- 4 En otra comunidad autónoma
- 5 Extranjero

B12. Año de llegada al municipio de residencia

Posibles valores:

- 2001 Año 2001 o anterior
- 2002–2018 2002–2018 (año a año)

### **3.4. Variables temáticas referidas a la persona**

Todas las variables temáticas procedentes de los ficheros administrativos integrados con la BD LPA que se han recogido en el fichero muestral de microdatos cuentan entre sus posibles valores con una categoría etiquetada con un 0 (cero) o una secuencia de ceros. En tal caso, se está indicando que la persona a la que se refiere el microdato no consta en el registro administrativo sectorial al que está asociado la variable. Por consiguiente, si se desea practicar una selección de las personas recogidas en el fichero muestral que, por ejemplo, sean receptoras de pensiones contributivas, bastará con seleccionar aquellas que cumplan la condición de que su valor en la variable P02 (o P01, o P03) sea distinto de 0.

#### **3.4.1. Información de pensiones contributivas**

P01. Número de prestaciones percibidas

P02 Tipología de la prestación

Posibles valores:

- 0 La persona no consta en el fichero de pensiones contributivas
- 1 Incapacidad
- 2 Jubilación
- 3 Viudedad
- 4 Orfandad
- 5 Favor familiar
- 6 Jubilación + Viudedad
- 7 Viudedad + Incapacidad
- 8 Otras combinaciones

P03. Importe mensual bruto acumulado

Posibles valores:

- 0 La persona no consta en el fichero de pensiones contributivas
- 1 Menos de 400 euros/mes
- 2 Entre 400 y 600 euros/mes
- 3 Entre 601 y 1.000 euros/mes
- 4 Entre 1.001 y 1.500 euros/mes
- 5 Más de 1.500 euros/mes

### 3.4.2. Información de afiliación a la Seguridad Social

#### A01. Número de afiliaciones

Posibles valores:

- 0 La persona no consta en el fichero de afiliaciones
- 1 1 afiliación
- 2 2 afiliaciones
- 3 3 afiliaciones
- 4 4 afiliaciones o más

#### A02. Régimen de afiliación a la Seguridad Social

Posibles valores:

- 000 La persona no consta en el fichero de afiliaciones
- 111 Régimen General excluidos Sistema Especial Agrario y Sistema Especial de Empleados del Hogar
- 112 Régimen General. Sistema Especial Agrario
- 113 Régimen General. Sistema Especial de Empleados del Hogar
- 121 Régimen Especial Trabajadores Autónomos
- 130 Régimen Especial del Mar
- 200 Pluriempleo
- 300 Pluriactividad

#### A03. Actividad económica del régimen o de la cuenta de cotización

Posibles valores:

- 0 La persona no consta en el fichero de afiliaciones
- A Agricultura, ganadería, silvicultura y pesca
- B Industrias extractivas
- C Industria manufacturera
- D Suministro de energía eléctrica, gas, vapor y aire acondicionado
- E Suministro de agua, actividades de saneamiento, gestión de residuos y descontaminación
- F Construcción
- G Comercio al por mayor y al por menor; reparación de vehículos de motor y motocicletas
- H Transporte y almacenamiento
- I Hostelería
- J Información y comunicaciones
- K Actividades financieras y de seguros
- L Actividades inmobiliarias
- M Actividades profesionales, científicas y técnicas
- N Actividades administrativas y servicios auxiliares
- O Administración Pública y defensa; Seguridad Social obligatoria
- P Educación
- Q Actividades sanitarias y de servicios sociales
- R Actividades artísticas, recreativas y de entretenimiento
- S Otros servicios

- T Actividades de los hogares como empleadores de personal doméstico; actividades de los hogares como productores de bienes y servicios para uso propio
- U Actividades de organizaciones y organismos extraterritoriales
- 1 Todas en el sector agrario
- 2 Todas en el sector industria
- 3 Todas en el sector construcción
- 4 Todas en el sector servicios
- 5 Varios sectores

#### A04. Tipo de contrato

Posibles valores:

- 0000 La persona no consta en el fichero de afiliaciones
- U100 No consta tipo de contrato en régimen cuenta ajena
- U110 Indefinido, no consta la jornada
- U111 Indefinido, tiempo completo
- U112 Indefinido, tiempo parcial
- U120 Temporal, no consta jornada
- U121 Temporal, tiempo completo
- U122 Temporal, tiempo parcial
- U200 Cuenta ajena, no procede tipo de contrato
- P110 Todas las afiliaciones por cuenta ajena, al menos una con contrato indefinido
- P120 Todas las afiliaciones por cuenta ajena y todas con contratos temporales
- P130 Todas las afiliaciones por cuenta ajena, otra situación
- P210 Una relación laboral por cuenta propia y al menos un contrato indefinido
- P220 Una relación laboral por cuenta propia y otra/s por cuenta ajena todas temporales
- P230 Una relación laboral por cuenta propia y otra/s por cuenta ajena en otras situaciones distintas a las anteriores

### 3.4.3. Información de demandantes de empleo

#### D01. Clasificación de la demanda

Posibles valores:

- 0 La persona no consta en el fichero de demandantes
- 1 Demandantes de otros servicios
- 2 Demandantes ocupados
- 3 Otros demandantes excluidos del paro registrado
- 4 Trabajadores eventuales agrarios subsidiados
- 5 Otros demandantes no ocupados
- 6 Parados registrados

#### D02. Ocupación en la que el demandante desea trabajar preferentemente

Posibles valores:

- 00 La persona no consta en el fichero de demandantes
- 01 Directores y gerentes
- 02 Técnicos y profesionales científicos e intelectuales
- 03 Técnicos; profesionales de apoyo



- 04 Empleados contables, administrativos y otros empleados de oficina
- 05 Trabajadores de los servicios de restauración, personales, protección y vendedores
- 06 Trabajadores cualificados en el sector agrícola, ganadero, forestal y pesquero
- 07 Artesanos y trabajadores cualificados de las industrias manufactureras y la construcción (excepto operadores de instalaciones y maquinaria)
- 08 Operadores de instalaciones y maquinaria, y montadores
- 09 Ocupaciones elementales
- 10 Ocupaciones militares

D03. Tiempo que lleva inscrita la demanda

Posibles valores:

- 0 La persona no consta en el fichero de demandantes
- 1 Menos de un mes
- 2 De 1 mes a 6 meses
- 3 De más de 6 meses a 1 año
- 4 De más de 1 año

D04. Ámbito territorial de búsqueda de empleo

Posibles valores:

- 0 La persona no consta en el fichero de demandantes
- 1 Municipal
- 2 Supramunicipal
- 3 Provincial
- 4 Comunidad Autónoma
- 5 Estatal
- 6 Unión Europea
- 7 Otros

D05. Actividad económica en la que ha trabajado anteriormente

Posibles valores:

- 0 La persona no consta en el fichero de demandantes
- A Agricultura, ganadería, silvicultura y pesca
- B Industrias extractivas
- C Industria manufacturera
- D Suministro de energía eléctrica, gas, vapor y aire acondicionado
- E Suministro de agua, actividades de saneamiento, gestión de residuos y descontaminación
- F Construcción
- G Comercio al por mayor y al por menor; reparación de vehículos de motor y motocicletas
- H Transporte y almacenamiento
- I Hostelería
- J Información y comunicaciones
- K Actividades financieras y de seguros
- L Actividades inmobiliarias
- M Actividades profesionales, científicas y técnicas

N	Actividades administrativas y servicios auxiliares
O	Administración Pública y defensa; Seguridad Social obligatoria
P	Educación
Q	Actividades sanitarias y de servicios sociales
R	Actividades artísticas, recreativas y de entretenimiento
S	Otros servicios
T	Actividades de los hogares como empleadores de personal doméstico; actividades de los hogares como productores de bienes y servicios para uso propio
U	Actividades de organizaciones y organismos extraterritoriales
Z	Sin empleo anterior

#### **3.4.4. Información de titulares de MUFACE**

##### M01. Situación administrativa

Posibles valores:

0	La persona no consta en el fichero de mutualistas MUFACE
1	En excedencia o suspensión
2	En actividad
3	Pensionista de incapacidad permanente o jubilado

##### M02. Actividad económica en la que trabaja (o ha trabajado)

Posibles valores:

0	La persona no consta en el fichero de mutualistas MUFACE
H	Transporte y almacenamiento
O	Administración Pública y defensa; Seguridad Social obligatoria
P	Educación
Q	Actividades sanitarias y de servicios sociales

#### **4. Control del riesgo de revelación y preservación del secreto estadístico**

La difusión de los resultados de toda actividad estadística se enfrenta siempre a un conflicto entre el derecho a la información y el derecho a la intimidad. El dilema se concreta entre el nivel de detalle con el que difundir la información, como medida de la utilidad de la misma, y paralelamente la necesidad de preservar el secreto estadístico, o lo que es lo mismo, la identidad de los informantes. Dada la riqueza y detalle de la información contenida en la BDLPA, incrementada aún más tras su integración a nivel individual con registros administrativos sectoriales, este conflicto es especialmente manifiesto.

En este producto se ha optado por una estrategia de difusión que maximice el nivel de detalle de la información bruta suministrada a los usuarios con objeto de que éstos puedan realizar sus propios análisis, adecuados a sus objetivos e intereses. La principal limitación en cuanto al nivel de detalle y desagregación de las variables del fichero muestral de microdatos publicado está motivada por el hecho de que el incremento del detalle amplifica proporcionalmente los riesgos de revelación. Por lo que, en último extremo, el máximo nivel detalle ofrecido es una solución de compromiso entre la utilidad e interés de la información y el riesgo de revelación a él asociado.

Se han tomado dos medidas tendentes a garantizar la necesaria preservación del secreto estadístico en el fichero de microdatos con una muestra de registros individuales:

- Los registros han sido anonimizados mediante la eliminación de todas las variables que sirven como identificadores directos de las unidades individuales.
- Se han considerado las combinaciones clave de variables que pudieran dar lugar a la identificación indirecta de las unidades individuales y se ha seguido una estrategia de recodificación global de las categorías de las mismas, con objeto de reducir el nivel de detalle informativo de dichas combinaciones y así minorar aceptablemente su potencia identificativa.