

METODOLOGÍA DE LOS AFILIADOS A LA SEGURIDAD SOCIAL SEGÚN EL LUGAR DE RESIDENCIA

1. Introducción

Los registros de afiliados y cuentas de cotización inscritos en la Seguridad Social suponen una fuente fundamental para analizar la evolución del mercado laboral y de la economía en general. Sin embargo, los datos que publica mensualmente la Seguridad Social no utilizan un concepto unificado respecto a la desagregación geográfica de los afiliados: mientras que para los trabajadores por cuenta propia usa como referencia el domicilio de residencia del trabajador, para los trabajadores por cuenta ajena utiliza el domicilio de inscripción de la cuenta de cotización a la que se adscribe el afiliado.

En consecuencia, con esta nueva operación estadística, el *Institut d'Estadística de les Illes Balears* (Ibestat) pretende dar datos del número de afiliados a la Seguridad Social según el domicilio de residencia del afiliado, y saber, por tanto, cuantas personas residentes en un determinado municipio están afiliadas, independientemente de que trabajen en el mismo municipio o en cualquier otro de las Illes Balears.

2. Construcción del *fichero único* de afiliados a la Seguridad Social

Trimestralmente el Ibestat recibe los registros de afiliaciones y cuentas de cotización a la Seguridad Social. Estos registros presentan problemas inherentes a cualquier fuente administrativa. En concreto, en cuanto a la explotación estadística, destacan los problemas derivados del registro de la actividad económica y de los datos personales de los trabajadores.

Así pues, el Ibestat ha diseñado un procedimiento para construir un *fichero único* que elimina los problemas citados anteriormente y que servirá de fuente tanto para la explotación propia de los afiliados a la Seguridad Social como para satisfacer las peticiones de información de los usuarios.

2.1 Depuración del código de actividad económica

El registro de la actividad económica en los ficheros de la Seguridad Social presenta dos problemas de carácter estructural. En primer lugar, la presencia de registros en los que el código de la Clasificación nacional de actividades económicas (CNAE) no es completo, hecho que se solucionó a partir del 1 de enero de 2007 cuando la Seguridad Social introdujo una nueva tarifa de primas vinculada a la actividad económica tanto a empresas como a trabajadores por cuenta propia. Por otra parte, se observa la existencia, a lo largo de todo el horizonte temporal, de registros que contienen códigos de la CNAE erróneos.

Para solucionar estos problemas se ha desarrollado un procedimiento que identifica todos los registros que presentan códigos CNAE problemáticos y les imputa una actividad correcta. En concreto, el proceso valida todos los códigos registrados siguiendo tanto la CNAE-1993 como la CNAE-2009. En caso de que la actividad registrada no sea correcta se extrae, de manera iterativa, el último dígito de la actividad registrada hasta alcanzar un código real, siempre que el código resultante tenga como mínimo dos dígitos (dado que los resultados de la explotación del *fichero único* se cuadrarán con los resultados publicados por la Seguridad Social, que utilizan la CNAE a dos dígitos).¹ Cuando todos los códigos son correctos, aunque no sean completos, se recurre a un proceso de imputación de los dígitos restantes para alcanzar un código de actividad con todos sus dígitos.

En términos generales, este procedimiento utiliza la información futura de un mismo registro (ya sean afiliados o cuentas de cotización), dado que, como se ha indicado, a partir del 2007 todos los registros presentan la actividad económica completa. En concreto, se trata de identificar el mismo registro en diferentes momentos del tiempo y complementar la información de la actividad del registro más antiguo utilizando el registro más reciente, siempre que la actividad del registro más reciente sea igual a la actividad del más antiguo en relación con el número de dígitos que éste tenga.

En resumen, el proceso mantiene todas las actividades rellenas perfectamente y válidas mientras que depura el resto. En caso de que la actividad sea incorrecta se elimina progresivamente el último dígito hasta alcanzar un código válido y, siempre que tenga como mínimo dos dígitos, se intenta estimar el resto de dígitos mediante la información futura del mismo registro.

2.2. Cruce con el padrón municipal de habitantes (a partir del 2011)

Con la intención de obtener una base de datos actualizada y de ofrecer información sobre las características personales de los trabajadores afiliados, a partir del 2011 los datos trimestrales de afiliados se cruzan con los datos del padrón municipal de habitantes. De este procedimiento se obtiene un conjunto de variables personales del trabajador, en concreto: la fecha de nacimiento, el sexo, la nacionalidad y los datos correspondientes al domicilio de residencia (municipio y distrito y sección censal).

2.2a Período de referencia del cruce

Para el cruce se utiliza el padrón *provisional* de habitantes, información que semestralmente (con día de referencia 1 de enero y 1 de julio, respectivamente) recibe el Ibestat. A diferencia de los datos *definitivos* del padrón, la base de datos provisional incluye las variables de identificación del individuo (como el DNI/NIE, el nombre, etc.), información que es crucial para realizar el cruce.

En referencia al período temporal utilizado en el cruce de ambos ficheros, se ha establecido que los datos de afiliados del primer trimestre del año t se crucen con el padrón a 1 de enero

¹ Véase el punto 4.

del año t , mientras que para el segundo y el tercer trimestres se utilice el padrón a 1 de julio del año t , y que, finalmente, para el cuarto trimestre se use el padrón a 1 de enero del año $t+1$.

Tabla 1. Ficheros para cruzar según el período

<i>Fichero de afiliados a la Seguridad Social</i>	<i>Padrón municipal de habitantes</i>
Primer trimestre del año t (datos a 31/03)	1 de enero del año t
Segundo trimestre del año t (datos a 30/06)	1 de julio del año t
Tercer trimestre del año t (datos a 30/09)	1 de julio del año t
Cuarto trimestre del año t (datos a 31/12)	1 de enero del año $t+1$

2.2b Validación del DNI/NIE

Dado que la variable fundamental para realizar el cruce de ficheros es el número de identificación de la persona (ya sea el DNI o el NIE), en primer lugar se ha llevado a cabo un proceso de validación de este identificador en ambos ficheros.

En el caso del fichero de afiliados, la validación consiste en reproducir el dígito de control del DNI/NIE y comprobar que coincide con el registrado. Si no hay coincidencia, el registro no se incluye en el procedimiento del cruce con el padrón. Por su parte, en la base de datos del padrón municipal, se produce tanto la validación del dígito de control del DNI/NIE como su imputación en caso de que el registro no lo incorpore (hecho que es habitual en el fichero provisional del padrón). Nuevamente, todos los registros que no pasen la validación no forman parte del proceso de cruce.

Por otro lado, se eliminan todos los registros, tanto del fichero de afiliados como del padrón municipal, que presenten números de identificación duplicados.

2.2c Niveles de cruce

Una vez validados los identificadores se inicia el proceso de cruce, que se estructura por niveles que se diferencian por el grado de exigencia en las condiciones de cruce. De esta manera se identifican cuatro niveles a lo largo del proceso:

- Nivel 1: se cruzan los registros con DNI/NIE correctos y no duplicados, coincidentes tanto con el DNI/NIE como con la fecha de nacimiento completa.
- Nivel 2: se cruzan los registros con DNI/NIE correctos y no duplicados, coincidentes tanto con el DNI/NIE como con el año y el mes de nacimiento.
- Nivel 3: se cruzan los registros con DNI/NIE correctos y no duplicados, coincidentes tanto con el DNI/NIE como con el año de nacimiento.
- Nivel 4: se cruzan los registros con DNI/NIE correctos y no duplicados, coincidentes sólo con el DNI/NIE.

La tabla 2 recoge el porcentaje de registros de afiliados cruzados con el padrón municipal para cada nivel a lo largo de todos los trimestres del 2011.

Tabla 2. Porcentaje de cruce según el nivel

<i>Período</i>	<i>Nivel 1</i>	<i>Nivel 2</i>	<i>Nivel 3</i>	<i>Nivel 4</i>
2011T1	90,1	2,1	1,3	0,9
2011T2	88,3	2,0	1,3	0,8
2011T3	89,3	2,0	1,3	0,8
2011T4	91,7	2,0	1,3	0,8

2.2d Tratamiento de los registros no cruzados

En cuanto a todos los registros no cruzados a lo largo del proceso anterior, como también todos los que contengan DNI/NIE incorrectos o duplicados, se les estima, si es posible, el valor de las variables personales que debería haber suministrado el padrón.

En concreto, para la fecha de nacimiento y el sexo se utiliza la información que proviene del mismo fichero de afiliados. De manera similar, el municipio de residencia se rellena con la información del domicilio de residencia declarado por el trabajador (información que consta en el mismo fichero de afiliados), si bien en el supuesto de que radique fuera de las Illes Balears se considera *sin descripción*. De la misma manera, se consideran siempre *sin descripción* tanto el distrito como la sección censal del domicilio de residencia del trabajador. Finalmente, para la estimación de la nacionalidad del individuo, se utiliza la información sobre el país de nacionalidad que incorpora el fichero de afiliados, sin embargo, al utilizar clasificaciones de países diferentes (ISO vs. propia del INE) los códigos se convierten utilizando una tabla de conversión entre ambas clasificaciones.

2.2e Resultados del cruce

La tabla 3 recoge el porcentaje de registros de afiliaciones cruzados con el padrón municipal de habitantes para cada trimestre del 2011. Como se observa, mediante la metodología expuesta anteriormente, se consiguen cruzar cerca del 94 % de los registros, siendo este porcentaje ligeramente inferior en los trimestres de temporada alta.

Tabla 3. Resultados del cruce para el año 2011

<i>Período</i>	<i>Registros totales</i>	<i>Registros cruzados</i>	<i>Porcentaje de cruce</i>
2011T1	389.918	368.374	94,47 %
2011T2	465.796	430.425	92,41 %
2011T3	431.403	403.427	93,52 %
2011T4	349.669	334.973	95,80 %
Total 2011	1.636.786	1.537.199	93,92 %

Por otro lado, a partir de los registros cruzados, es posible analizar las diferencias entre las variables personales que consten en cada registro. En este sentido, la tabla 4 recoge la diferencia porcentual de las variables *municipio de residencia*, *sexo* y *fecha de nacimiento*.

Tabla 4. Diferencia porcentual entre variables personales de los registros cruzados

<i>Período</i>	<i>Municipio de residencia</i>	<i>Sexo</i>	<i>Fecha de nacimiento</i>
2011T1	16,16	1,30	4,58
2011T2	17,00	1,33	4,43
2011T3	16,35	1,31	4,46
2011T4	16,02	1,26	4,31

3. Consistencia con los datos de la Seguridad Social

Con el objetivo de mantener un *dato único* entre la Seguridad Social y el Ibestat, el fichero resultante del cruce anterior se modifica para incorporar todas las correcciones que introduce la Seguridad Social a la hora de publicar sus resultados.

Además, dado que en algunos casos la fecha de extracción de los datos de la Seguridad Social no coincide con la fecha de referencia de sus publicaciones, el Ibestat lleva a cabo un procedimiento de encuadre que garantiza el mantenimiento del *dato único*.

4. Explotación y encuadre de los resultados

La explotación de los datos del *fichero único de afiliados* a la Seguridad Social que hace el Ibestat complementa la información que ya se publica trimestralmente en el web del Instituto. En concreto, la diferencia fundamental reside en la utilización de las variables personales de los trabajadores obtenidas del padrón, en lugar de utilizar las que recoge la Seguridad Social. Pero, dado que la explotación de los microdatos no cuadra siempre con los resultados publicados por la Seguridad Social y que la voluntad del Ibestat es mantener un *dato único*, se ha optado por aplicar un sistema de encuadre que asegura que los resultados que publica el Instituto sobre las variables laborales (régimen de afiliación y actividad económica a dos dígitos de la CNAE) sean los mismos que los presentados por la Seguridad Social.

En concreto, el procedimiento implantado utiliza la distribución porcentual obtenida de la explotación de los microdatos aplicándolos a los totales publicados por la Seguridad Social (por período, régimen de afiliación y actividad económica). De esta manera, se consigue obtener información de las variables personales de los afiliados extraídas del padrón sin incumplir el propósito del *dato único* de afiliaciones.

En cuanto a la explotación propia, el Ibestat publicará trimestralmente un conjunto de siete tablas con la información de las afiliaciones de alta del último día del trimestre. Las tablas publicadas por el Instituto son las siguientes:

- I302039_0001: afiliados a la Seguridad Social según el lugar de residencia por período, sección de actividad (CNAE-2009), isla y municipio de residencia.
- I302039_0002: afiliados a la Seguridad Social según el lugar de residencia por período, sección y división de actividad (CNAE-2009), isla de residencia.
- I302039_0003: afiliados a la Seguridad Social según el lugar de residencia por período y régimen de afiliación.
- I302039_0004: afiliados a la Seguridad Social según el lugar de residencia por período, isla y municipio de residencia y grupo de edad quinquenal.

- I302039_0005: afiliados a la Seguridad Social según el lugar de residencia por período, isla y municipio de residencia y sexo.
- I302039_0006: afiliados a la Seguridad Social según el lugar de residencia por período, isla de residencia y continente de nacionalidad.
- I302039_0007: afiliados a la Seguridad Social según el lugar de residencia por período, isla y municipio de residencia y zona de nacionalidad.

5. Comparación de los resultados con la Seguridad Social

Como se ha indicado al principio, los datos publicados por la Seguridad Social de los afiliados por municipio incluyen como municipio de referencia el del domicilio de residencia y el de la cuenta de cotización, dependiendo del tipo de trabajador. En cambio, mediante el procedimiento expuesto anteriormente, el Ibestat facilita el número de afiliados a la Seguridad Social por municipio de residencia del trabajador. La tabla siguiente recoge las diferencias entre ambas fuentes para el primer tercer del 2012; a modo de ejemplo, en el municipio de Esporles las cifras de afiliados según el lugar de residencia (fuente Ibestat) suponen un aumento del 88,52% sobre los afiliados publicados por la Seguridad Social.

Tabla 5. Afiliados por municipio (Seguridad Social vs. Ibestat), tercer trimestre de 2012

<i>Isla y municipio</i>	<i>Seguridad Social</i>	<i>Ibestat</i>	<i>Variación %</i>
ILLES BALEARS	440.460	440.460	0,00
MALLORCA	348.550	332.581	-4,58
07001 Alaró	964	1.874	94,42
07003 Alcúdia	8.821	8.391	-4,88
07004 Algaida	896	1.908	112,90
07005 Andratx	3.702	4.638	25,27
07006 Artà	1.805	2.849	57,85
07007 Banyalbufar	164	230	40,20
07008 Binissalem	1.682	2.838	68,72
07009 Búger	155	352	126,93
07010 Bunyola	1.559	2.501	60,41
07011 Calvià	23.085	22.230	-3,71
07012 Campanet	483	852	76,34
07013 Campos	2.692	3.898	44,81
07014 Capdepera	5.147	5.402	4,95
07016 Consell	885	1.483	67,62
07017 Costitx	140	386	175,67
07018 Deià	527	249	-52,76
07019 Escorca	185	132	-28,74
07020 Esporles	976	1.894	94,10
07021 Estellencs	102	129	26,38
07022 Felanitx	5.085	6.708	31,91
07025 Fornalutx	164	266	62,14
07027 Inca	8.248	10.694	29,66
07028 Lloret de Vistalegre	206	416	101,80
07029 Lloseta	1.173	1.998	70,34
07030 Llubi	383	751	95,98
07031 Llucmajor	9.649	14.461	49,87
07033 Manacor	14.020	15.146	8,03
07034 Mancor de la Vall	157	513	226,67
07035 Maria de la Salut	390	668	71,41
07036 Marratxí	14.475	14.678	1,40
07038 Montuïri	560	983	75,45

07039 Muro	3.373	2.791	-17,25
07040 Palma	193.518	148.638	-23,19
07041 Petra	702	1.007	43,44
07042 Pollença	5.630	6.202	10,16
07043 Porreres	1.451	1.993	37,33
07044 Pobla (Sa)	2.817	4.569	62,18
07045 Puigpunyent	351	710	102,22
07047 Sencelles	467	1.161	148,52
07049 Sant Joan	386	700	81,38
07051 Sant Llorenç des Cardassar	2.559	3.459	35,15
07053 Santa Eugènia	287	649	126,06
07055 Santa Margalida	5.502	4.987	-9,37
07056 Santa Maria del Camí	1.527	2.527	65,48
07057 Santanyí	6.058	5.554	-8,31
07058 Selva	664	1.393	109,75
07059 Salines (Ses)	1.849	2.057	11,22
07060 Sineu	747	1.276	70,83
07061 Sóller	3.911	5.651	44,49
07062 Son Servera	6.868	5.598	-18,49
07063 Valldemossa	578	797	37,82
07065 Vilafranca de Bonany	551	1.012	83,70
07901 Ariany	274	337	22,99
MENORCA	30.346	33.731	11,15
07002 Alaior	2.968	3.436	15,76
07015 Ciutadella de Menorca	9.119	10.563	15,83
07023 Ferreries	1.511	1.954	29,33
07032 Maó	11.656	10.247	-12,09
07037 Mercadal (Es)	1.865	1.721	-7,71
07052 Sant Lluís	1.735	2.469	42,29
07064 Castell (Es)	1.201	2.787	132,03
07902 Migjorn Gran (Es)	291	555	90,68
EIVISSA	55.523	57.947	4,37
07026 Eivissa	26.138	21.596	-17,38
07046 Sant Antoni de Portmany	7.656	9.877	29,01
07048 Sant Josep de sa Talaia	8.822	10.583	19,97
07050 Sant Joan de Labritja	1.965	2.065	5,11
07054 Santa Eulàlia del Riu	10.942	13.826	26,35
FORMENTERA	4.168	4.761	14,22
07024 Formentera	4.168	4.761	14,22
SIN DESCRIPCIÓN	1.873	11.440	510,78