Population Census

Methodology

Provisional version

Contents

INE. National Statistics Institute

1   Introduction

After the publication of the 2021 Population and Housing Censuses[1], a paradigm shift in demographic statistics took place, given that the methodology used in the 2021 census - which for the first time was based entirely on the processing and exploitation of dozens of official administrative records - allows for the process to be replicated each year thus providing annual population censuses.

Thanks to this new methodology, it will be possible to have updated information every year, in the case of the population, and every three or four years in the case of households, instead of every ten years, as was the case with the decennial censuses that have been carried out until now. In addition, these new annual censuses are prepared with information collected entirely from official public administration records, without the need to survey any household.

The **"Population Census"** thus emerges as a new statistical operation of a structural nature and on an annual basis, with code **8892** in the National Statistical Plan 2021-2024, and with code **30281** in the Inventory of Statistical Operations. This operation will be separate from the housing census, which will have a lower frequency.

As a general rule, from 2024, the first results of the Population censuses will be published in December each year, for the reference date of 1 January of that same year.

However, in this first edition, the first results of the censuses on 1 January 2022 and 2023 are published together.

The "First results", as they are called, offer the official figures of the resident population in Spain on 1 January in great territorial detail (up to the census section), together with their first demographic characteristics.

In the coming months, this demographic information will be completed with variables related to education and employment, among others, from the exploitation of thematic administrative files provided by several different organisations.

2   Objectives

Among the objectives that the Population Census operation aims to cover, it is worth highlighting the traditional goals of a census:

a) Population count. The Population Census determines the official population figures of Spain and all its administrative divisions (the autonomous communities, provinces, municipalities as well as census districts and sections) on 1 January of each year.

b) Knowing the structure of the population. The Population Census not only counts the number of inhabitants but investigates, for each of them, a series of demographic,

economic and social characteristics and provides a snapshot of the population structure.

c) To meet international statistical needs. International organisations request information from countries to compile their demographic and social statistics, with censuses being one of the main sources used.

Specifically at the EU level, at the time of this first publication of results, the legislative framework governing demographic statistics is undergoing renewal.

For several years now, the relevant bodies of the European statistical system have been working on a new regulation that consolidates the demand for statistics on stocks and demographic flows, which are currently dispersed in three different regulations[1] (one covers annual migration statistics, another demographic phenomena and annual stocks and a third is that of the decennial censuses).

At the time of preparing this methodological document, there is as yet no conclusive proposal, but there is already a sufficiently mature draft, which would place most of the current demands of the census on an annual basis. In general, the new European population regulation is expected to call for **greater territorial detail** (with particular emphasis on georeferencing), **greater frequency** and **greater timeliness** of information on the population and associated demographic phenomena.

This need for more frequent and faster demographic information, which already existed in society, has been increased by the events of recent years (Brexit, the COVID-19 pandemic, the war in Ukraine, etc.). With the current information requirements, it has been proven that it took too long to find out the demographic impact they had.

Annual censuses will make it possible to respond to these more demanding population data requirements.

But beyond the traditional objectives covered by census information, it should be noted that this new strategy based on registries to construct census information constitutes a comprehensive renewal in demographic information:

d) The availability of annual census information has given rise to a new system of demographic statistics that are consistent with one another, making it possible to eliminate the duplication of demographic information.

Traditionally, after the publication of each ten-year census, a statistical population series would start. Specifically, starting with the 2011 Population and Housing Censuses, and beginning with 1 January 2012, the Population Figures series was published, estimated by means of an accounting of demographic events (births, deaths, internal and external migrations and acquisitions of Spanish nationality by the foreign population).

Additionally, the publication of a new census every ten years entailed a correction of the population data series of the entire previous decade until the previous census (the traditionally called intercensal estimates). With the availability of annual

---

[1] At present, demographic statistics are regulated by three Eurostat regulations of the European Parliament and of the Council: Regulation 1260/2013 on European demographic statistics, Regulation 862/2007 on Community statistics in the field of migration and international protection, and Regulation 763/2008 on population and housing censuses.

censuses, this system of decennial intercensal corrections will no longer be necessary.

Until now, two sets of population statistics were published each year: those derived from the exploitation of the Municipal Register and what we could call "statistics", which were constructed starting with the 2011 census. More specifically, two sets of population stock statistics existed side by side:

– The half-yearly Population Figures (with reference dates of 1 January and 1 July).

– The Continuous Census Statistics, annually (with a reference date of 1 January).

From 2023, both operations disappeared, replaced by the afore-mentioned annual population censuses and the rollout in May 2023 of a new short-term population stock statistics, the **Continuous Population Statistics**, fully consistent with the annual population censuses.

In this way, not only census information is available as of 1 January of each year with the maximum territorial detail, but also quarterly population estimates, very swiftly (with a time lag of a maximum of 45 days with respect to the reference date, compared to the 5 or 6 months of the old Population Figures) and fully consistent with the census figures, although with less territorial detail (up to the province/island).

This new operation also provides information on the population living in family homes and the number and size of households.

This responds to the demand by users, both outside and within the INE, to consolidate the population figures in a single consistent population series at all geographical levels, dispelling doubts about which batch of figures to use.

The creation of this new system of demographic statistics based on annual population censuses also affects migration statistics.

The former operations: Statistics of Migrations and Statistics of Residential Variations, both based on the counting and statistical treatment of certain types of residential changes registered in the continuous register, have been eliminated.

Instead, all recorded variations that may involve a change of residence between two census times and, at the same time, are consistent with every two consecutive censuses, will be treated. To respond to these characteristics, the **Statistics on Migrations and Changes of Residence** will be published for the first time at the end of 2023, replacing the previous two.

This new set of statistics will include migrations, both external and within the country, according to the definitions of international recommendations, and disaggregated by a higher territorial level. It will also allow, in the future, the study of intra-municipal changes of residence, for example between the boroughs of a city.

e) Annual census information enables the enrichment of other operations

Census files have a great potential for exploitation in themselves, due to the amount of information they contain to characterise the population. But they can also be cross-referenced at the individual level with other statistical operations. This has already been done for years in the operations of the Natural Movement of the Population. Thus, based on the census files, characteristics such as the level of education, occupation, relationship with activity or marital status are added to the parents of

newborn children, or of deaths. This potential to exploit individual characteristics will be exploited in more operations, such as the new Statistics on Migrations and Changes of Residence, and will make it possible to find out the nature of migratory movements in a qualitative, and not only quantitative, way.

3   Concepts and definitions

Resident population: persons, of any nationality, who have established their "habitual residence" in the national territory.

Primary residence: The place where a person normally spends their daily periods of rest, regardless of temporary absences for leisure travel, holidays, visits to family and friends, business, medical treatment or religious pilgrimage. Nonetheless, it must be noted that only the following will be considered regular residents in a territorial district:

– Those who, according to the above definition, have regularly resided therein for a continuous period of at least twelve months.

– Those who, according to the above definition, had established their regular residence therein less than twelve months before, but intended to stay there for at least one year.

When the circumstances described above cannot be determined, the "regular residence" will be understood to be the place of legal or registered residence.

Sex: Sex refers to the biological sex of the person.

Age: The age in completed years refers to the number of birth anniversaries on the reference date, i.e. the age on the last birthday.

Nationality: The special legal bond that unites an individual with their State, acquired through birth, naturalization, declaration, choice, marriage and other formulas in accordance with national law. It refers to the nationality on the reference date, classified as Spanish provided that the individual has that nationality, while they may also have others.

Place of birth: Place where a person was born, i.e. the mother's habitual place of residence at the time of birth, determined on the date the data are collected.

Completed level of education: The level of education or training attained by a person is defined as "the highest level that the individual has completed" and "is measured from the highest level programme that he or she has successfully completed, usually validated by the relevant certification".

Main current studies: The "Main Education" is provided, which is generally the General Regime education (Compulsory Secondary Education [E.S.O], Baccalaureate, Vocational Education [F.P.], university studies, etc.) of a higher level that the person is studying (if he or she is studying more than one).

Type of studies in progress: Field of a student's major university-level courses of study in progress.

Place of study: A student's registered place of study in relation to the student's usual place of residence.

Relationship with the economic activity: Current situation of the population aged 16 and over with respect to economic activity, based on a reference period of a recent week, which is usually the week prior to the census reference date.

4 The population census

4.1 THE CENSUS POPULATION FIGURE: THE SIGNS OF LIFE METHOD

The prime objective of censuses is to count the number of inhabitants residing in a territory on a reference date, which in the case of annual population censuses is 1 January of each year.

The census population figure should be considered as the **official population figure at all territorial levels.**

The method of estimating the resident population in the annual censuses is the one already used for the 2021 population census: the "signs of life" or "signs of presence" method.

These methods essentially consist of starting from a population register, in our case the Municipal Register, and evaluating the plausibility of each person's residence based on their presence in other registers, so that for each individual it is decided whether to count them as a resident or not, in a more precise way than the official administrative accounting of the Municipal Register.

Although the signs of life method improves the quality of the census population figure, **the key element of the process is the existence of a population register**, **the Padrón - Municipal Register**, whose use as the "skeleton" of the population census has major advantages over other possible alternative population registers (tax registers, health cards or General Directorate of the Police file). All residents are required to register in the register of the municipality where they live most of the year, including a population that at first glance may seem difficult to access, such as people without legal residence or registered in non-conventional housing. In addition, since the implementation of the Continuous Register more than 25 years ago, with the INE as the coordinating body of the more than 8,000 municipal registers, the Municipal Register has matured a great deal as a population register, establishing, among others, expiry procedures and periodic verification of the residence of foreign nationals that make the Register a reliable and accurate updated population register.

The theoretical case and the practical limitations of a signs of life method

In theory, in today's society it is almost impossible to live without leaving a trace in some kind of register, so if a statistical office had unlimited access to such information, which is not the case, it would seem that it could be determined quite accurately for each person whether or not they reside in the country (or in a particular place).

If we focus strictly on official administrative records, that is, publicly owned (not considering access to social networks, use of credit cards or bank accounts, mobile telephony or other privately owned systems) the situation changes radically. Even so, there could be administrative records that offer very relevant information about

presence: public transport tickets, medical consultations, obtaining medicines, information from city and town halls, traffic, social services, etc.

In practice, to date, there are hardly any files of this type that can be used for a population census and many technical, data protection or public opinion barriers would have to be overcome before a population census could make use of these sources to determine the residence of people, even for statistical purposes as is the case.

But there are certain administrative files at our disposal that are part of the census project and that can help to determine the number of residents somewhat more accurately than if we limit ourselves to the purely administrative count that the census figure yields. But not only that, if we do not use them, we will be making important errors and inaccuracies not only in the number of residents but also in some variables that the census collects, such as those related to economic activity.

Pre-2021 Census trials of signs of life methods

The work began with the pre-census files of 2016 and 2018 and used a procedure, with several alternatives, that was unrealistic because it would force too much of the population to be discounted as there were no signs of life for it.

The Municipal Register was taken as a basic element of the structure of the pre-census file and from it a procedure was developed to carry out a population count by cross-referencing it with other files.

Then, the signs of presence of the people of the base in the available official administrative records were analysed, as well as the census movements in the months after the reference date.

Among the **administrative information** [i.e. from official agencies] used to determine these signs of a person's presence in Spain, it is worth highlighting:

– Information from the Tax Agency, Provincial Treasuries and Territorial Treasury of Navarre.

It takes into account whether the person appears as a Personal Income Tax filer for the previous year or the existence of withholdings and payments on account, distinguishing between those made for various items (income from work, income from economic activities, benefits received, minimum non-contributory allowances for the unemployed, etc.).

Information contained in the file of insured persons and beneficiaries registered on 1 January of each year in the Health Insurance Database (BADAS) provided by the INSS (National Social Security Institute).

In this file, insured persons are classified into insurance groups according to their status (workers, pensioners, etc.). There are specific groups that include insured persons from other countries and their beneficiaries who are residing in Spain whose health expenses are borne by the country of insurance and for whom a form has been received guaranteeing that the country in question provides their health coverage in Spain. It is considered that in these cases residence in Spain is highly likely.

There is another group that includes people who lost their initial insurance for various reasons, for example a worker who loses their job, and at the end of the extension

period they have not found another form of insurance. These people are less likely to be residing in Spain.

In general, however, the BADAS file is of little use for signs of life, since its nature is to guarantee health care and therefore it has less tendency to erase population.

– Sources related to the labour market

The information from the files of job seekers on the last day of the months of November and December of the year prior to the reference date, provided by the State Public Employment Service, is analysed, taking into account that the person meets the conditions to be considered unemployed for the purposes of the SEPE (Spanish Public Employment Service).

The file of Social Security members who were registered during the week prior to 1 January of each year, provided by the General Treasury of the Social Security, as well as the file of benefit recipients on 1 January of each year of the National Social Security Institute, is also used.

Mutual society members, both insured and beneficiaries, are also considered on 1 January of the main mutual societies (MUFACE, MUGEJU; ISFAS).

In general, the presence of a person, with certain exceptions, in these files can be taken as a sign of presence for the process of calculating the population.

– Other complementary sources

Information from the Central Register of Foreign Nationals, such as expiry dates and proof of residence, dates of application, granting and refusal of residence permits, etc.

Initially, all these indicators of presence in the various official administrative files are cross-referenced at the individual level, but this results in around 10% of the population either having no signs of life, or having them but they are not very significant. The second occurs if we count the records that cross-reference with the BADAS file, due to their tendency not to erase population.

An inactive person, who lives with another person who has a low income, and therefore does not have to file a tax return, would not appear in any of the previous records, but that is not a sufficient reason to discount them in a population count (we would be erasing several million people). In general, the population not found in signs of life is mostly female, increases with age and affects foreign nationals more than Spaniards.

The application of signs of life at the individual level also has an important implication in terms of households. For example, if a person does not appear in signs of life and is erased, this person disappears from a household whose composition is altered and can even give rise to meaningless households (minor children living alone).

That is why the signs of life analysis was undertaken in a second revision, in terms of households. A range of alternatives were tested on the basis that just one member showed signs of life and the entire household was considered confirmed.

Even with this approach, around 1.5% of the population (which means about 700,000 people) have no signs of presence. The analysis of this population by geographical distribution and sociodemographic characteristics suggested that the population would

be eliminated due to defects in the cross-referencing of files rather than due to true signs of life.

The signs of life method Used in annual population censuses

The method used since the 2021 census **is applied exclusively to the foreign population**, since it is only for them that we have the expiry or ex officio verification processes that the signs of life method aims to correct.

In addition, it applies at the household level, so that only households in which all household members are unequivocally affected or the resulting household is not altered in composition are no longer counted, for example, households in which there are no kinship relationships between family members.

The municipal register currently operates by applying procedures for the renewal of the foreign population and periodic verification of residence that allow the file to be kept reasonably updated. It is somewhat similar to what other countries do with population registers or how registers such as the DNI (national ID card) work, which require renewal from time to time.

The foreign population, for the purposes of register management, is divided into two groups, "ENCSARP" and "non-ENCSARP". Non-EU foreign nationals without a permanent residence permit (abbreviated ENCSARP) are people who reside in Spain without having Spanish nationality or that of any other State of the European Union and who do not have a permanent residence permit issued by the General Directorate of the Police. Since 2005, these people have had to go to their respective city or town halls to renew their registration in the municipal register every two years. If they do not do so, a process is set in motion that ultimately results in their removal from the municipal register due to expiry (called BBC in the terminology of the register).

The rest of the foreign population is known as "NON-ENCSARP". These foreign citizens residing in Spain are not obliged to renew their registration in the municipal register, but since 2008 a gradual procedure has been put in place whereby local councils must periodically confirm their residence. At first it was done only for a percentage of them, but now it is checked for all of them. This check must be carried out for all foreign nationals in this group who have not made any movements in a period of five or two years, depending on whether they are registered in the Central Register of Foreign nationals or not. If it is not possible to verify that the person continues to reside in the municipality, the city/town council in question will launch a procedure that will culminate in an ex officio deregistration due to improper registration (BII).

In short, this procedure means that for the foreign population as a whole it can be analysed whether the inhabitant's registration has reached its "expiry" date (strictly speaking, only the registration of the ENCSARP expires, but we can speak of expiry for the rest when those two or five years have elapsed). In principle, we could stop counting as residents all foreign nationals in the census database whose registration has expired.

If we analyse a cohort of foreign nationals who enter Spain in a given year, [following] the current process and assume that all of them expire after two years, we would have the situation schematically illustrated in the following diagram:

Diagram of the foreign population count according to the expiry date of their registration
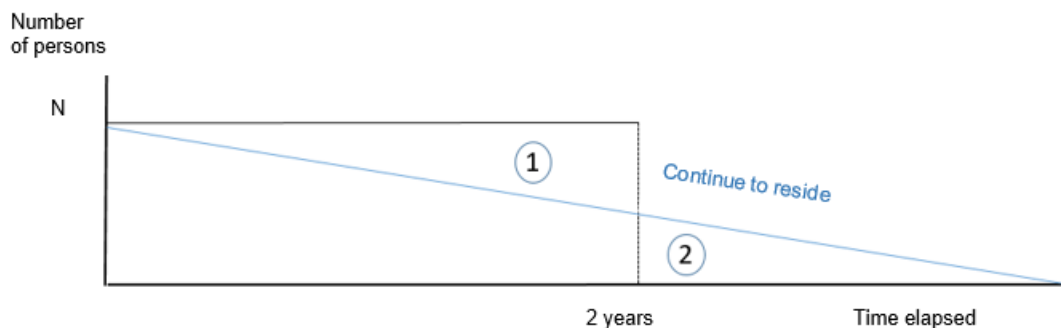
According to this method, the N people in the cohort would count as long as their registration had not expired, and all of them would automatically stop being counted after those two years.

But the immigrating population leaves the country more as shown by the descending blue line. Triangle 2 contains the population that continues to live in Spain but whose registration has expired. If we cross-reference this population with the files of signs of life (labour, tax) we find foreign nationals who would initially be "expired" but who, because they present sufficient signs of presence, are "retrieved" and counted in the census population figure.

Triangle 1 contains the population that no longer resides in Spain but whose registration has not yet expired, so they continue to count as inhabitants. The signs of life method applied does not have an immediate analogy with the one used for group 1, because for the reasons mentioned above, the non-presence in these records does not seem to suffice to stop counting a registered person. But it can be done more clearly for some of them. The decision taken since the 2021 census is to apply signs of life only to a part of foreign nationals, those who did appear the previous year in the signs of life files. Thus, if a foreign person appeared in the pre-census file on 1 January 2020 as a member of a household or as an unemployed person or pensioner, student or in any other situation that denotes his or her presence, but on 1 January 2021 no longer appears in any of these files, we assume that the person has left the country, even if their registration has not yet expired.

Once this has been done, in a first phase, the homes resulting from the application of these signs of life must be analysed, and with this a second phase is applied that corrects the previous one.

Specifically, for children under 16 years of age, given that, in general, they are not present in the official administrative registers used in the signs of life procedure (except for the group of 15-year-old students who are enrolled in an official education), it is considered a sign of presence that there is at least one adult with signs of presence at the address at which they are registered. In other words, minors who, after applying the procedure of signs of presence, are registered in a home without adults, would no longer be counted.

Conversely, minors who, although initially having their registration expired, reside in households where at least one adult shows signs of presence in registers, would be counted in the census population.

In the method of signs of presence applied in the 2021 Census, the aforementioned employment-related records (INSS, TGSS, SEPE, Mutual Societies), some specific BADAS groups for which residency is guaranteed and, exceptionally, education-related records (enrolled in university or non-university education) have been used. Although tax information was also available, it was decided not to use it in the method. The reason is that tax information is received one year after the census reference date, and the objective from 2021 onwards was to obtain census population figures as of 1 January of each year that are available within the current year. In order for these census population figures to be consistent and form a series on an annual basis from 2021, we need to base the signs of life method on the same sources every year. For this reason, it has been decided to use the most rapidly available sources of labour and to discard the tax ones.

The population over the 100 year mark

As with any population registry, the census accumulates records that can be considered "remnants" of a procedure that cannot be perfect. This is the case of people who died, even decades ago and who were never deleted. They must be Spaniards, given that there are expiry processes for foreign nationals. It is proven every year that there are very few cases, but they do exist. There are several reasons that can lead to this, but fundamentally they are errors in the death certificate that prevent the deregistration of a registered person.

In that case, the record of the deceased person is kept in the register and "ages" indefinitely. These records come to light when they reach very advanced ages, when the number of effectives is already very low. This is why exhaustive control processes are put in place for the population aged 100 and over (called "centenarians").

Obviously, this problem of residual records that refer to the deceased population but are not deleted from the file occurs not only in centenarians but also in the entire census file; however, it is quite undetectable and only surfaces at those older ages. We must recognise the limitation that this entails and that it can mean a few thousand records, but its impact is very small within much larger population groups and that is why the effort is concentrated on the population aged 100 and over.

To mitigate these defects, the birth dates of the population aged 100 or over are first checked by cross-referencing them with the DNI database and carrying out a small number of purges.

Apart from this, a process of correction of the population aged 100 or over is applied so that any census record of more than 99 years of age that cannot be linked to any INSS (pensioners) or tax record is deleted from the census file, because it is considered highly unlikely that it refers to a living person.

Other population not included in the Municipal Register: ages 0 and 1

Another element that differentiates the census figure from the one collected in the census is the population of very young ages, mainly under 2 years old. Due to the administrative method used in the census, a small part of the births are added somewhat late to the register. These births have been detected and the information upon registration in the civil registry as well as the statistical birth bulletin is available for them. Obviously, those records are added to the census count. In the specific case of 2021, it involved some 22,000 children aged 0 or 1 years, almost all of them of Spanish nationality.

4.2 DEMOGRAPHIC VARIABLES: GENDER, AGE, NATIONALITY, PLACE OF BIRTH

In this first publication of results, only the main demographic variables of the population are offered.

This information is obtained mainly from the data contained in the register.

A small number of records in the country of birth and nationality are necessary, affecting a residual number of records, applying an imputation process when the information is not found in other official administrative records.

A process to detect possible inconsistencies in the sex variable is also carried out, cross-checking with sources from the Ministry of the Interior, as well as the date of birth.

In addition, since the place of birth is defined as the mother's place of habitual residence at the time of birth, it has been necessary to search for this information in the census, keeping, when possible, the child's place of residence (which must match that of the mother) instead of the place of birth.

4.3 EDUCATION-RELATED VARIABLES: LEVEL OF EDUCATION ACHIEVED, MAIN STUDIES IN PROGRESS, PLACE OF STUDY

After the first publication of the population census, the variables of an educational nature are provided for the population aged 15 years and over.

This information comes from numerous official administrative sources with personal information on educational topics, which are combined in the census to obtain the educational variables of people aged 15 and over.

Breakdown of the classification of education-related variables:

The level of education attained is broken down on the basis of an adaptation of the CNED (National Education Classification)-A[1] with a maximum detail of twelve categories:

– Illiterate persons

– Incomplete primary education

– Primary education

– First stage of secondary education and the like

– Second stage of secondary education, with general orientation

– Second stage of secondary education, with professional orientation

– Non-higher post-secondary education

– Vocational training, visual arts and design and sports training of a high level and equivalent

– University degrees of up to 240 ECTS credits, university diplomas and equivalent qualifications

---

[1] https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177034&menu=ultiDatos&idp=1254735976614

INE. National Statistics Institute

- University degrees with more than 240 ECTS credits and equivalent qualifications

- Master's degrees, specialities in Health Sciences for the intern system and similar

- University doctorate

The CNED-A encompasses formal education programs, both current and non-current. It should be noted that we do not have full coverage of all the categories of the CNED-A, since, for example, there is no information available on university-specific degrees (CNED-A 52, 63, 75). In addition, there is partial coverage of non-tertiary post-secondary education (CNED-A 41), as well as lower levels of professional certificates (CNED-A 23, 24).

Most of these more problematic categories are immersed in others of the adapted CNED-A (Higher Degree or University Degrees in the case of own degrees; first stage of secondary education in the case of vocational education and training certificates 1 and 2).

The **main courses of study** in progress are broken down on the basis of an adaptation of the CNED (National Education Classification)-A[1] with a maximum detail of twelve categories:

- Primary education or lower

- First stage of secondary education and the like

- Compulsory Secondary Education (4th year)

- Baccalaureate-level education and similar

- Vocational, visual arts and design and sports training of a mid-level and equivalent

- Education in Official Language Schools

- Professional education in music, dance and similar

- Vocational training, visual arts and design and sports training of a high level and equivalent

- University degrees of up to 240 ECTS credits, university diplomas and equivalent

- University degrees with over 240 ECTS credits, bachelor's degrees and equivalent qualifications

- Master's degrees, specialities in Health Sciences for the intern system and similar

- Doctorate

Non-formal education programmes are not included.

The **type of education in progress** is broken down on the basis of the CNED-F1 with a maximum detail of ten categories:

- Education

- Arts and Humanities

- Social sciences, journalism and documentation

---

[1] https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177034&menu=ultiDatos&idp=1254735976614

- Business, administration and law

- Science

- IT

- Engineering, industry and construction

- Agriculture, livestock, forestry, fishing and veterinary

- Health and social services

- Services

Source of education-related variables:

The level of education attained is **the highest found among all the records of the person**, and these two circumstances occur at the same time:

- More often than not, one finds a person in more than one educational-themed source. This is the case for about 85% of people aged 15 and older.

- No record is at the same time complete, current, and provides information with the required disaggregation.

A good example of this dichotomy is the Municipal Register of residents as a source of education-related data. Every person who registers must fill out a census form (municipal register sheet) that allows them to provide the data according to the required CNED-A. However, not all of the population is listed at that level of detail.

The management rules for the Municipal Register allow municipalities to classify this variable in two ways. This means that information on school certificates is not available for all persons with the same level of detail. Thus, especially in old censuses, there are only four categories.

Approximately 75% of the population has the detailed level and for the remaining 25% there are only four possible aggregate levels of education. Additionally, this information may become obsolete unless the person changes their registration and fills in a new registration form.

In order to update the level of education, the register has been cross-referenced for several years with the register of Qualifications of the Ministry of Education and Vocational Training (hereinafter, the MEyFP), correcting the initial information in many cases. Therefore, the municipal register offers complete coverage, but requires further refinement.

The estimation of the level of education attained is carried out by updating the educational information of each person. The sources used are as follows:

- **Municipal Register.** The school or academic title is a registration variable that must be collected by municipalities in their municipal register. As such, it appears on the registration forms. Inclusion in the Register is related to the creation of the Electoral Census. Thus, in the ongoing management of the municipal register, this information is collected from the city and town councils, and is filtered using information received every six months by the Ministry of Education and Vocational Training (MEyFP).

- **Qualifications awarded by the Ministry of Education and Vocational Training.** Information is available from the files of non-university (LGE, LOGSE, LOE) and

university degrees and accreditation of foreign degrees as equivalent to university and non-university degrees, which the Ministry of Education and Vocational Training sends to the National Statistics Institute on an ongoing basis.

– **2001 Census**. For the nearly 34.5 million people 16 years of age or older who responded to the census questionnaire in 2001 with a reference date of 1 November 2001, there is information on the level of education achieved up to that date.

– **2011 Census**. For the nearly 3.5 million people 16 years of age or older who responded to the census questionnaire in 2011 with a reference date of 1 November 2011, there is information on the level of education achieved up to that date.

– **Job seekers from the State Public Employment Service (SEPE).** For persons enrolled in public employment services, information on the level of studies completed is available.

– **SEPE professional certificates**. A record exists of historical professionalism certificates that certify skills recognized by the educational Administration; this serves to validate the corresponding professional module or modules.

– **University graduates.** Historical information is available for university graduates for the 2010-2011 academic year until the academic year prior to the census.

– **Holders of non-university degrees.** Historical information is available for university graduates for the 2014-2015 academic year until the academic year 2020-2021, prior to the census.

– **Central Registry of Foreigners.** Information is available for foreign nationals with currently valid authorisation to study in Spain.

– **SEPE contracts in force.** Since 2001, information has been available that each person's last contract as of 1 January for the reference year and gives the worker's educational level.

– **Enrolled in the educational system in non-university education of the Ministry of Education and Vocational Training (MEyFP).** Available from the 2014-2015 academic year onwards.

– **Enrolled in university education through the Council of Universities.** All enrolments are available from the 2010-2011 academic year onwards.

All these sources are considered for the estimation of educational attainment. For people aged 16 and over who resided in Spain on 1 January of the previous census reference, their educational level is updated.

The coverage of the level of education attained is almost total. Around 1.5% of people aged 15 and over are attributed to their educational level, although for the vast majority of them there is information in one or more educational records, with only the desired detailed classification remaining to be specified (for example, whether it is a university degree of up to 240 ECTS credits or more).

In the case of information on **studies in progress**, the last two sources referring to enrolments are mainly used. The Central Registry of Foreign Nationals is also used to detect new students based on their residence permit for students, as well as loading the MEyFP Qualifications to a date after the census reference date to deduce enrolments.

Finally, the registered **place of study** is obtained from the same sources as the studies in progress, mainly from the files of university enrolments and non-university enrolments of the MEyFP. Also, the Central Registry of Foreign nationals contains information to know the province where most of the students detected in this way are studying. For about 92% of the students it has been possible to extract the municipality where they are studying and for an additional 0.4% the place of study is offered at the provincial level. This leaves a small remnant of about 8% of unassigned study space.

Remarks on the variable of studies in progress

In the publication of the annual censuses, it has been decided to **filter the studies in progress** of a small percentage of students (approximately 1%), whose enrolments are too low for their attained level of education. Enrolments in programmes of a lower level have been allowed in combination with studies achieved in controlled cases: for example, after confirming that they are enrolments in elementary education (music and dance, or adult education), enrolments in basic vocational training of pupils who have entered it despite not having completed lower secondary education, or certain enrolments in baccalaureate-level or further ("adult") education.

Nearly a thousand enrolments that were too high for the age of the student have also been purged.

These people have been assigned a specific category ("They are in education but there is no information about them").

In addition, it should be noted with regard to **studies in progress of the younger population** that their assignment has been resolved in a small percentage of cases by imputation and cannot yet be considered fully resolved.

On the one hand, the files received from the Autonomous Communities, although they have been improving drastically in recent years, are not yet fully exhaustive. Thus, data are not available for some private schools or for some particular courses. This means that it cannot be guaranteed that 100% of the student body will be available.

On the other hand, it is necessary to link the information that comes from the educational system with the population census file and the link is not perfect either. Especially for foreign nationals, the identifier (DNI, NIE) is not always available to improve the link, but people appear in the enrolment files with their passport number. Thus, in the files of students enrolled in non-university education, there are about 2% of unlinked enrolments. And, on the other hand, the forenames and surnames do not appear in the files of the university system, which would allow a slightly higher percentage of links to be achieved than the current one, which means that just over 4% of university enrolments are not linked.

In addition, the possibility of people who are enrolled in Spanish universities but reside abroad and are doing distance learning must be considered. In total, there is a small remnant of around 3% of the files of studies in progress that cannot be linked to any person in the population file, which in part may be due to degrees obtained or studies taken by non-residents but in other cases it is the typical problem of the lack of a link due to lack of good identifiers (absence of ID or forenames and surnames for many records).

It also happens the other way around: there are people who must surely be studying but for whom no link can be found with any enrolment. In Spain, in general, education is

compulsory until the age of 16. This means that students who are 15 years old on the census date must necessarily be in some kind of education. But due to the limitations in the files of the educational system and their cross-referencing with the census population file, in particular there are 4% of 15-year-olds for whom it cannot be ascertained what course of study they are pursuing, even though by law they should be doing so. These people have been assigned a specific category ("Following a course of study but there is no information about them").

Remarks on the variable "place of study":

It should be borne in mind that the main use given to the variables "Place of work" and "Place of study" is to determine origin-destination mobility matrices. The variable "origin", which would be the municipality of residence according to the Register, in a small number of cases may not reflect reality and give rise to apparently strange origin-destination flows. But it must also be considered that the context of remote working and distance learning, increased during 2020 by the pandemic, may increase these situations that at first sight appear strange. Attempting to identify errors and make corrections could be risky and it has been preferred not to make corrections.

4.4 LABOUR-RELATED VARIABLES: RELATIONSHIP WITH ECONOMIC ACTIVITY AND VARIABLES RELATED TO THE EMPLOYED (PROFESSIONAL SITUATION, ECONOMIC ACTIVITY, PLACE OF WORK, OCCUPATION)

After the first publication of the population census, the relationship with economic activity for the population aged 16 and over, based on the week prior to the reference date of the census, is provided.

This information comes from numerous official administrative sources with personal information on labour-related topics, which are combined in the census to obtain the statistical variable.

For people classified as "employed", information about their professional situation, economic activity, workplace and profession is offered.

Sources of the variable "Relation to economic activity":

The following sources have been taken into account:

– Employment-related sources:

- **Workers who contribute to Social Securit**y who were registered (General Treasury of Social Security) during the week prior to 1 January of each year.

- **Spanish Public Employment Service (SEPE) job seekers**.

- **Civil servants** who do not contribute to Social Security and are members of one of these three **mutual insurance companies**: MUFACE, MUGEJU and ISFAS.

- **Persons registered with the Social Security System** (INSS) who received a pension, and the reason for this pension (retirement, disability, orphanhood, widowhood).

– Educational information on students enrolled in the education system, both for university and non-university education (see previous section).

– Tax information from form 190 of the Tax Agency and similar forms from the Provincial Treasuries.

– Information contained in the file of insured persons and beneficiaries registered on 1 January of each year in the Health Insurance Database (BADAS) provided by the INSS (National Social Security Institute).

For every person aged 16 and over, information is sought from employment and educational sources in the first instance. Additional information of a tax nature is sought for people who are not in them, or for those from whom only a situation of inactivity other than retirement or disability can be deduced (for example, recipients of orphan's or widow's benefits).

Since so many sources are taken into account, it is normal for there to be different coexisting situations, which lead to conflict for each person. For example, a person who registered as a job seeker and in turn, enrolled on a course of study or training course.

In spite of this, and in contrast to the previous section on educational variables, it is most common for a person to only appear in an official administrative source in relation to employment. Or that, if the person appears in several registers, the information is the same. In general, each administrative source has almost complete coverage of a particular employment situation:

– The majority (94.2%) of people employed in the census are included in the file of members of the General Treasury of the Social Security Agency. The percentage increases if mutual societies are taken into account (98.7%).

– In the case of the unemployed, they are included in the SEPE's job seekers file (96.3%).

– The majority of people receiving retirement or disability benefits come from the National Social Security Institute (90.5% and 95.5% respectively). The percentage increases if mutual societies are taken into account (97.3% and 95.9%, respectively).

– The majority of students (93.7%) are enrolled in university and non-university education.

Remarks on the reference date:

It is important to consider the reference date for this data. The Regulation establishes the following:

'Current activity status' is a person's present situation with respect to economic activity, based on a one-week reference period, which may be an indicated, recent, fixed, calendar week, the full calendar week or the seven days preceding the census.

Since the reference date of the census is 1 January of each year, it would then be a matter of obtaining the employment status of the population during the previous days. For this reason, a specific request has been made to the General Treasury of the Social Security in an attempt to reflect the membership situation during the last full week prior to the reference date, and the information has been constructed taking all the people recorded as active during that week as employed.

Rules for assigning the relationship with labour activity:

In the event that the information in a person's employment and education records does not match, deterministic decision rules are applied to classify them into a single situation. This affects about three million people.

A series of priority rules have been established based on the recommendations of the United Nations and the European Regulation for Censuses:

– If the person appears as employed in any source of data, they are considered to be an employee.

– If the person receives several benefits, the retirement benefit takes precedence over the rest for those over 65 years of age (55 in the case of the unemployed).

These rules are also applied depending on age:

– If the person is over 65 years of age and cannot be found in any administrative source, they are considered inactive.

– If the person is exactly 15 years old, and is not found in any other administrative source, he or she is considered a student.

– In the case of persons under 65 years of age who have not been found in any official administrative source, they are placed in the category of other situations of inactivity. The percentage of non-assigned is approximately 9.5% of the population aged 16 and over. The fact that a person is listed as unassigned in a category like this should not appear strange. In fact, it stands to reason that many of the inactive ones do not appear in those records.

No corrections, such as age-related refinements, are made to the variable resulting from applying these rules.

No se hacen correcciones, ni por ejemplo depuraciones en función de la edad, sobre la variable resultante de aplicar estas reglas.

Sources of the variables related to the employed:

The following sources have been taken into account:

– **Workers who contribute to Social Securit**y who were registered (General Treasury of Social Security) during the week prior to 1 January of each year.

– **Civil servants** who do not contribute to Social Security and are members of one of these three **mutual insurance companies**: MUFACE, MUGEJU and ISFAS.

– Information regarding the last current contract of each person as of 1 January of each year in the Public State Employment Service (SEPE).

– CBR (Central Business Register).

– Withholdings for income from work as an employee (Spanish Tax Authority [AEAT] and Provincial Treasuries).

– Information on previous censuses.

For each employed person, their working conditions are evaluated according to the information in the sources for each year compared to the previous year. If it is considered that they remain the same, the economic activity and occupation of the

previous reference is maintained. It can be said that it 'inherits' information from the previous reference. Otherwise, it proceeds according to the methodology of the Population and Dwellings Censuses 2021[1].

The evaluation of the employed persons paying Social Security contributions for each year with respect to the previous year is made according to their professional situation, their National Classification of Economic Activities (NCEA) activity, their type of employment relationship and their contribution group, current and previous, and the date of their contract. The evaluation of the rest of the persons employed is made according to their professional situation, their NCEA activity, and the date of their contract.

4.5   VARIABLES RELATED TO MIGRATIONS: YEARS OF ARRIVAL, PLACES OF ORIGIN AND PREVIOUS RESIDENCE

Years of arrival:

Information is provided regarding the following years of arrival:

– Year of arrival in Spain.

– Year of arrival in the autonomous community.

– Year of arrival in the province.

– Year of arrival in the municipality.

For all of them, it refers to the year of arrival of the last stay in that geographical area. For example, if a person has resided in Spain from birth until 2017, then resided abroad for three years and returned to Spain in 2020 and has resided here since then, their year of arrival in Spain would be 2020.

Autonomous community and province of origin:

The autonomous community and province of origin are calculated taking into account the migratory movement by means of which the person arrived in the autonomous community or province of census residence, which may be different, since a person may change province within the same autonomous community.

For example, if on 1 January 2022 a person resides in Ciudad Real and that same person had a migration from Toledo to Ciudad Real during the year 2021, then for the reference date 1 January 2022 their province of origin will be Toledo. In contrast, to find his or her autonomous community of origin, it is necessary to go further back in time. If the same person, before residing in Toledo, resided in Seville, then their autonomous community of origin would be Andalusia.

Previous residence:

The above residence variable takes into account the following values:

– They have always resided in the same municipality.

– Other municipality in the same province.

– Another province in the same autonomous community.

---

[1] https://ine.es/censos2021/censos2021_meto.pdf

- Another autonomous community.

- Was temporarily abroad.

- No record of previous residence, but has always resided in Spain.

- No record of previous residence, and has not always resided in Spain.

For its calculation, current residence is compared with previous residence involving at least one change of municipality.

Sources of migration variables:

The migratory variables are obtained through the information from the previous annual population census, taking as a starting point the 2021 census together with the information offered by the Migration and Change of Residence Statistics (MCRS)[1] of the previous year. For example, to calculate the migration variables of the census with reference date 1 January 2022, the census with reference date 1 January 2021 is used together with the Migration and Change of Residence Statistics for the year 2021.

For those persons who were already present in the previous census and who have not had any migration during the year in a specific geographical area, the migration information they had in that census is 'inherited'. For those persons who did have some migratory movement during the year, the migratory information collected in the Migration and Change of Residence Statistics is also taken into account.

For example, if a person during the year 2022 has a migration from Valladolid to Zamora, the variables of year of arrival in Spain and year of arrival in the autonomous community for the 2023 census will be the same as for the 2022 census, while the year of arrival in the province and municipality will be the year 2022.

5 Links of official administrative records with the municipal register

The foundation of a census based on official administrative records is the linking or cross-referencing of people's recorded information with residents. For this reason, requests for data from official administrative records bring the personal information that identifies each person: the type of TIN, in general the national ID for Spaniards and the foreign national ID for foreign nationals, forename and surname (as separate variables), full date of birth, sex, and, if available, nationality, place of residence or any variable which, although it does not identify the person by itself, does serve as a validation parameter for the link with the census.

There are several noteworthy aspects about the link of people with the census:

- Several linking strategies are required to consider that an administrative source has been successfully linked to the standard.

- Most of the sources used in the census involve a minimum of half a dozen cross-references. In general, as many strategies per source have been devised as it has been necessary to achieve a link rate of close to 100%.

- Each strategy employed consists of two phases:

[1] https://www.ine.es/metodologia/meto_emcr_2023.pdf

- Cross-referencing or linking one or more variables identifying the person (between the source and the register).

- Validation of the cross-reference or verification that the cross-referencing carried out does not link up due to excess.

In addition to these phases, there is often a duplicate removal phase. For example, in a cross-reference by forename and surname, if there are several people in the census who have the same name, the link must be made with only one of them. The decision of which one to link to would be based on the rest of the variables identifying the person, in this example it could be made based on the date of birth.

– In particular, a 100% close link percentage would not be satisfied with the national ID/foreign national ID alone; nor could the cross-references be validated, which is another crucial aspect of the process.

In the case of foreign nationals, the quality of personal data beyond their national ID number, passport, or identity card, is a good indicator of quality when it comes to linking to the census. This is because different official administrative sources may have been updated at different points in time.

A more general example of the difficulties involved in receiving a source with few identifying variables has been described in section 4.3 – *Remarks on the variable of studies in progress* (link of university enrolments).

Link percentages:

The greater the quality of the identifiers, as well as the rest of the variables that identify the person, the higher the percentage of links between official administrative records and the census.

It has been accepted that the percentage must be close to 100% in order to be considered satisfactory in order to consider the link between a register and a census satisfactory, but this statement is full of nuances.

In reality, it depends on the nature of the source. For example, a source such as the SEPE contract history, because it contains information from several years, will have a lower ideal link percentage. Another example is the file of Social Security members, which despite being a register with excellent data quality, it will never be possible to reach 100% of members linked to the register. Even if the process were perfect, registered members who are not residents in Spain will always be (correctly) left unlinked.

If the history of contracts is excluded, the percentage of links from sources on employment-related issues is 98.5%. Among them, civil servants' mutual societies are close to the maximum with 99.6% linked; the sources from the Social Security agency, which are much more numerous and contain a richer set of scenarios of employment situations and nationalities, link 98.9% (Treasury) and 97.8% (INSS) thanks to having a high quality of identifying data.

With regard to university and non-university enrolment files, as mentioned in section 4.3 - *Remarks on the variable of studies in progress*, the percentage of links is 97.0%, although 1.7% of them are **probabilistic links**. Probabilistic means a method of linking in which a registered person is not found in the population register, but is assigned as a proxy for another resident with whom he or she shares demographic characteristics

(such as age, sex, nationality or place of residence). For statistical purposes, this enrolment counts as linked for all purposes on the census estimate, although it is purely an estimate that we make for the person on the municipal register of residents, to whom we match up enrolment information through a donor.

Data preparation and linking strategies:

Before linking to the register, the personal information received from the registries is standardised. During this phase, identifiers, dates of birth and, above all, the literals of forenames and surnames are standardised. The fact of having two surnames or compound names is an advantage when it comes to making links. This entails, in turn, a long time to prepare and debug these identification variables.

Thanks to this phase, patterns have been found that are exploited during the cross-referencing phase. For example, the same type of error in date composition (such as mixing up the month and day) is common. It is also common to find forenames and surnames out of order. In the validations of each cross-referencing strategy, there are tools such as the calculation of alphanumeric distances or distances in dates of birth that take into account this type of scenario, common in many records. Scores can also be assigned regarding the commonness or uncommonness of a surname, or even deducing with some margin of error whether or not a person is a foreign national based on their first and last name.

However, not all sources provide information in the same way. For example, in some tax records, first and last names are provided together in a single variable. This poses an added difficulty in the strategies of crossing these sources, trying to mitigate them specifically.

Once the recorded information received has been filtered, the most common cross-referencing strategies are:

1. Links by identifier, with and without a control digit, with the letter of foreign national; Validation by distances in the date of birth and similarities in forename(s) and surname(s).

2. Links by same forename(s), surname(s) (ordered or not) and date of birth; validation by punctuation of surname(s).

3. Links due to identical forename(s) and surname(s)(ordered or not); validation by distances in date of birth.

4. Links by identifier (partial), without control digits or foreign letter, and date of birth; validation by similarities in forename(s) and surname(s).

5. It is also linked by identifier (partial) and validated by distances in date of birth and similarities in forename(s) and surname(s).

6. Links by similarities of forename(s), surname(s) and date of birth; validation by identical forename(s) and surname(s), or by punctuation of surname(s).

The use of variables such as gender or nationality has been simplified to summarise the most commonly used strategies, since date of birth or forename(s) and surname(s) are most commonly used for both links and validations. For example, if there are records without date of birth information or incomplete forename(s) and surname(s).

6   Dissemination of results

The first results of the annual Population censuses are published in December each year, for the reference date of 1 January of that same year. They offer the official figures of the resident population in Spain in great territorial detail (up to the census section), together with their first demographic characteristics.

In the first edition of the annual Population Censuses, in December 2023, the first results of the censuses on 1 January 2022 and 2023 were published together.

In April 2024, the demographic information of the population as of 1 January 2022 was completed with variables related to education and employment, among others, from the exploitation of official thematic administrative files provided by several different bodies.

In September 2024, the Annual Population Census 2022 information is completed with variables relating to employed persons, such as occupation, professional situation (salaried, self-employed, etc.), activity of the establishment and municipality of work.

On the other hand, information on the years of arrival in the different territorial areas and on previous residence has been added to the Annual Population Censuses 2023 and 2022. An extensive set of tables is published for the different geographical areas, up to the census section, which will contain the population counts classified by the main demographic, migratory, educational and labour-related variables.

They will be accompanied by thematic maps, mainly at the municipal level and census sections, for the main indicators.

INE. National Statistics Institute