Introduction

This user guide provides essential information to inform the use, analysis and interpretation of the 2011 Census Microdata Teaching File. This, anonymised, random sample of census records provides a useful resource for users of census outputs to analyse census data in a way that is not possible using the standard census tables.

www.ons.gov.uk/census

Office for
National Statistics

Table of Contents

1 Introduction

As with previous censuses, it is envisaged that a variety of microdata products will be released from the 2011 Censuses across the UK. The 2011 Census Microdata products provide an alternative source of data to the standard 2011 Census releases, which consist of tables containing counts of how many people in one area have a certain attribute or attributes.

In contrast, microdata contain information from the individual respondents which have been treated to protect the confidentiality of the respondents. Rather than data being delivered as tables of counts, data are stored in the same way as if the data had been collected by conducting an anonymous sample survey. Only a sample of cases is available and individual records contain information on a limited number of topics. The Teaching File is a microdata product with a relatively small sample and low level of detail that is freely available for anyone to download from the ONS website.

Different microdata products are being developed that provide more detail and hence more utility, but have more restrictions on use given the increased risk to confidentiality. To this end, the census offices are continuing to work alongside key users to develop and agree the specifications and access arrangements for those products, with the aim of further releases later in 2014.

1.1 Purpose of the Microdata Teaching File

The primary purpose of the Microdata Teaching File is as an educational tool to:
• encourage wider use of census data by providing a way of examining census data beyond the standard tables;
• provide an introduction to the detail, metadata and data formats included in microdata products in order to give users the skills and information necessary to make use of the more detailed products to come;
• assist with the teaching of statistics and geography at GCSE and higher levels.

1.2 Scope of the Microdata Teaching File

ONS is responsible for carrying out the census in England and Wales. Simultaneous but separate censuses took place in Scotland and Northern Ireland. There were run by the National Records of Scotland (NRS) and the Northern Ireland Statistics and Research Agency (NISRA) respectively. The Teaching File released by ONS covers the countries of England and Wales only. A similar product for Northern Ireland is available from NISRA. NRS have plans to produce a similar product for Scotland.

2 Confidentiality and protection of personal data

Preserving the confidentiality of personal information provided by the public on their census questionnaires remains a top priority for the census. (See more detail on data confidentiality in the 2011 Census outputs). The Microdata Teaching File contains a 1% sample of people with a small number of their characteristics. Additional measures have been taken to ensure that individuals cannot be identified:

• No personal identifiers (name, address, date of birth) have been included in the Microdata Teaching File

• Potentially disclosive output variables have been either completely removed or have been aggregated to reduce the level of detail available for each record.
In particular, geographic information is limited
to region (e.g. North West or London).

3 About the Microdata Teaching File

3.1 What does the Teaching File contain?

The Microdata Teaching File, or dataset, consists of a random sample of 1% of people in the 2011 Census output database for England and Wales. This includes people classed as both usual residents and short-term residents. Specifically, the dataset includes:
• Records on 569,741 individuals
• For each individual, information is available on 18 separate characteristics, or variables, with varying degrees of information for each variable. See the complete list of variables and classifications.

Further details on how these variables were derived from the 2011 Census questionnaire, including definitions and derivations, can be found in the 2011 Census variable and classification information document.

The Teaching File includes data from the complete England and Wales population, which includes usual residents, short-term residents and students living away from home during term-time (categorised in the 'population base' variable).

The main population base for published statistical tables from the 2011 Census is the usual resident population as at census day, 27 March 2011. For 2011 Census purposes, a usual resident of the UK is anyone who, on census day, was in the UK and had stayed or intended to stay in the UK for a period of 12 months or more, or had a permanent UK address and was outside the UK and intended to be outside the UK for less than 12 months.

In the Teaching File, of the 569,741 individuals in the dataset, 561,040 (98.5%) are classified as usual residents. If the data from the Teaching File are not filtered to include only usual residents, results may differ from published statistical tables. Many of the variables included in the dataset, however, contain values only for usual residents.

Students and schoolchildren in full-time education studying away from the family home were counted as usually resident at their term-time address. Basic demographic information only (name, sex, age, marital status and relationship) was collected at their non-term time ('home' or vacation) address. In the Teaching File, information collected at their non-term time address is available by filtering the Population Base variable to include only 'students living away from home during term-time'. Datasets - including both 'students living away from home during term-time' and 'usual residents' - could thus include duplicate information on basic demographic variables for some individuals. Therefore give careful consideration when choosing the appropriate population base for data exploration, and when comparing findings from the Teaching File with published reports.

3.2 How was the sample drawn?

3.2.1 Sample size

Sample sizes have been chosen to be consistent with statistical disclosure control considerations and draft user specifications. The number of records that are within the sample and unique within the census database was measured as a proportion of the number of records that are unique within the sample. The level for this proportion was set in order to determine sufficient uncertainty. The sample composition chosen will remain confidential, as was the case for the 2001 Census microdata.

3.2.2 Stratification

Stratification enables the characteristics of a sample to be proportionally representative of the population by dividing the population into strata based on key characteristics. Random samples taken from each stratum are then pooled to form the final sample.
The Teaching File sample is stratified by census output area within Local Authority. This method ensures good representation of data, the sample is more evenly spread, and is consistent with the user requirement for a multipurpose product that can be used for a wide variety of analyses. It also controls against extreme sample selection, ensuring for instance that an entire output area is not selected at random.

3.3 Limitations of the Microdata Teaching File

The Microdata Teaching file is a 1% random sample of people in the census. It is therefore subject to sampling error, which arises whenever variable estimates are based on a sample rather than a full census of the population. Sampling error is the difference between the estimates derived from a sample and the true population values.

Although the Teaching File sample is broadly representative when compared with similar univariate distributions from the entire census population (see comparisons with published statistics), care should be taken when generalising findings from any sample to a wider population, with due consideration to the reliability of results.
Given the relatively small sample size and the limited detail on each of the characteristics it may not always be the most appropriate dataset for analysis. For example:
• It is always more appropriate to use a published table that covers the entire population where that is available (see the complete list of all available tables).
• It may be more appropriate to use a more detailed classification (either available from a more detailed microdata product or a standard census table). For example 'age' in the Teaching File is given in broad age bands, one of which is 16-24 years of age. If differences in a variable of interest are perceived likely between those aged under 18 and those aged 18 and over, a dataset with a single year age breakdown may be more appropriate.

3.4 The 2011 Census

Full information on how the 2011 Census in England and Wales was conducted is available on the 2011 Census page on the ONS website http://www.ons.gov.uk/ons/guide-method/census/2011/index.html.
The 2011 Census is the most complete available source of information on the population. However, despite efforts to reach everyone and obtain the most accurate information possible, no census is perfect and some people are inevitably missed. Further information on how the 2011 Census was conducted and the treatment of missing data is available in the Quality and Methods section of the 2011 Census User Guide.

3.5 How should you cite the data?

The Open Government Licence, which applies to the teaching file, allows unrestricted use of government data as long as the source is acknowledged.

Crown copyright
All material on the Office for National Statistics (ONS) website is subject to Crown Copyright protection unless otherwise indicated.

Reproducing ONS material
Under the terms of the Open Government Licence (OGL) and UK Government Licensing Framework, anyone wishing to use or re-use ONS material, whether commercially or privately, may do so freely without a specific application for a licence, subject to the conditions of the OGL and the Framework. These new arrangements replace the previous Click-Use and Value Added Licences.
Users reproducing ONS content without adaptation should include a source accreditation to ONS: Source: Office for National Statistics licensed under the Open Government Licence v.1.0.
Users reproducing ONS content which is adapted should include a source accreditation to ONS: Adapted from data from the Office for National Statistics licensed under the Open Government Licence v.1.0.

4 Other Census products and getting more information

Further information from the 2011 Census on output products already published or planned for publication in the future can be found on the 2011 Census web pages.
For further information about census results please contact Census Customer Services.
Census ad hoc tables contain combinations of data that are not available in standard publications. They have been requested by the media or other user. Once created, ad hoc tables are published for all to use. Ad hoc tables will be constrained by the data that are available at each stage, by the similarity to what ONS plans to release, and by statistical disclosure control. Tables can be pre-ordered from Census Customer Services.

Need further help

Census Customer Services
Office for National Statistics
Segensworth Road
Titchfield
Fareham
Hampshire
PO15 5RR

census.customerservices@ons.gsi.gov.uk

Tel : 01329 444972
Fax : 01633 652981

Office for
National Statistics