

Census 2001: 10% Sample of unit records (Version 1)

**Thank you for choosing the Census 2001 10% sample of unit records.
We hope that you will find this product useful.**

1. GENERAL INFORMATION

This file sets out the information that you will need to access the data provided.

In order to process and analyse the data users should be in possession of appropriate software that can handle very large datasets. Users can contact Stats SA for advice in this regard if necessary.

The files essential for accessing the data are provided in text format. Other files with relevant information have been provided in Word 2000 format. If you do not have software appropriate for this format, please contact Stats SA and paper copies will be forwarded to you. Alternatively, this and additional documentation is available on the Stats SA website situated at www.statssa.gov.za

2. CONTENTS OF CDs

2.1 Data

This directory contains the data files in zipped format:

- Households.zip
- Persons.zip
- Mortality.zip
- Household imputation flags.zip
- Person imputation flags.zip
- Geography.zip

2.2 Metadata

This directory contains metadata for all the variables in the following files:

- Introduction.doc
- Households.doc
- Persons.doc
- Mortality.doc
- Imputation flags for households and persons.doc
- Geography.doc

2.3 Code lists

This directory contains code lists for:

- Country of birth and citizenship
- Religion
- Occupation
- Industry

All code lists are contained in the metadata on the Stats SA website. These four code lists are supplied separately for the convenience of users.

Users of the variables on migration and place of work should consult the main-place or sub-place code list on the website.

2.4 Questionnaires

This directory contains the following questionnaires in .pdf format

- Questionnaire A (for persons in households)
- Questionnaire B (for persons in institutions)
- Questionnaire C (for institutions)

2.5 Record layouts

This directory contains the record layouts of the files described in 2.1 above.

2.6 Definitions

This directory contains the concepts and definitions used in the data.

2.7 Adjustment factors

This directory contains an Excel file with four worksheets showing the adjustment factors for persons and households on municipality and provincial level, which can be used to calculate the universe.

If required, standard errors (SE) for each variable can be calculated by Stats SA.

3. REQUIREMENTS

The following minimum hard drive space is required:

- | | |
|------------|-------|
| • Data | 122Mb |
| • Metadata | 12Mb |
| • Other | 1Mb |

4. DESIGN OF THE SAMPLE

This sample is a 10% unit level sample drawn from Census 2001 as follows:

4.1 Households:

- A 10% sample of households in housing units, and
- A 10% sample of collective living quarters (both institutional and non-institutional) and the homeless.

4.2 Persons:

- A sample consisting of all persons in the households and collective living quarters, and the homeless, drawn for the samples described above in 4.1.

4.3 Mortality

- A sample consisting of all mortality information for the households in housing units drawn in the 10% sample of households.

5. WEIGHTING FACTORS

Both the 10% household and person sample files contain a weight variable. This weight variable is the adjustment factor for undercount (for households or persons as appropriate) multiplied by 10 to inflate

the 10% samples to the relevant population. In the person records, aggregated totals of sparsely populated codes, such as very old ages, might differ substantially from real totals due to sampling fluctuations – no scaling of the weights was done. In the household records aggregated totals will be approximately equal to real totals. Mortality was not adjusted for undercount and therefore there is no weight variable.

6. STRATIFICATION AND ORDERING OF THE RECORDS

The census household records were implicitly stratified according to municipality, geographic type and EA number. The latter is a unique eight-digit census Enumerator Area number.

The following geographic types were used:

- Urban formal
- Urban informal
- Tribal
- Rural formal

7. VARIABLES INCLUDED IN THE 10% SAMPLE

All variables as per the questionnaire are included in the 10% sample, as well as derived variables and imputation flags.

EA numbers are excluded to preserve confidentiality.

Geographic type is excluded from the final sample. Instead two additional geographical variables are supplied, namely:

- Urban and rural – Census '96 classification
- Size and density of locality

8. GEOGRAPHY

The South African geographical structure for the 10% sample consists of the following geographical entities, which fit into different geographical hierarchical levels:

South Africa
Province
District council (DC - Category C) or Metropolitan area (Category A)
Magisterial districts (MD)
Local municipality (Category B), or District management area (DMA)

While the structure is intended to be hierarchical, South Africa's geography has cross-boundary entities, which complicate the picture. For example, there are eight municipalities which lie across provincial boundary lines. Users are advised to bear this in mind when choosing the appropriate hierarchy. For example, for the City of Tshwane, which lies in two provinces, one would not use the provincial hierarchy.

Due to the existence of cross-boundary entities there are five distinct geographical hierarchies.

9 MERGING THE DATASETS

Number of records in the datasets :

- Households : 948 592

- Persons : 3 725 655
- Mortality : 36 267
- Household imputation flags : 948 592
- Person imputation flags : 3 725 655
- Geography : 948 592

Serial number is a common variable in all files listed in 2.1 above. This variable together with the variable *Person number* can be used to merge Persons with their relevant Imputation flags.

Serial number can also be used to merge all files with the different geographical hierarchies in the file Geography.zip.

The variable *Type of living quarters (comprehensive)* is included in both the Households and Persons files to assist with the analysis of the data.

10. INTERPRETING THE DATA

10.1 Confidentiality

In order to preserve confidentiality the lowest geographical level that unit records can be linked to is municipality.

As further assurance of the confidentiality of the data, municipalities with 200 or fewer households are logically grouped with adjacent municipalities.

The following municipalities are grouped:

Code	Grouped with
193	114
292	218
491	415
591	592
691	605

10.2 Extract from the *Report of the Census Sub-Committee to the South African Statistics Council on Census 2001*

“Preliminary investigations indicate that the 2001 census probably resulted in:

- an underestimate of the number of children below age five*
- an over-estimate of the number of teenagers aged between 10 and 20
- an underestimate of the number of men relative to the number of women*
- an underestimate of the number in the white population
- higher than expected numbers aged 80 and older, in the African population
- an underestimate of the number of foreign-born, since some identified themselves incorrectly as being South African-born
- age misstatement in the range 60-74
- an overestimate of the extent of unemployment
- an underestimate of those who were employed for only a few hours per week
- an underestimate of household income
- an overestimate the number of paternal orphans and the number of fathers missing from the household.

* This is a common feature of censuses, particularly in developing countries.

In addition:

- Scanning problems caused some births to be recorded in the wrong province. The number of cases is relatively small and should not lead to too much distortion for most purposes for

which these data are used; however, it does produce obviously erroneous results when one tries to estimate the extent of inter-provincial migration of those born since the previous census.

- The fertility data (numbers of children ever born, children surviving) are problematic.

For further details of these investigations see the full report of the *Census Sub-Committee*.”

11. COPYRIGHT NOTICE AND DISCLAIMER

© Copyright, Statistics South Africa, 2003.

The information products and services of Stats SA are protected in terms of the Copyright Act, 1978 (Act 98 of 1978). As the State President is the holder of State copyright, all organs of State enjoy unhindered use of the Department’s information products and services, without a need for further permission to copy in terms of that copyright.

Where a copy of the information is made available to any third party outside the State, the third party must be made aware of the existence of State copyright and ownership of the information by the State.

The State (through Statistics SA) retains the full ownership of its information, products and services at all times; access to information does not give ownership of the information to the client. The use of any data is subject to acknowledgement of Stats SA as the supplier and owner of copyright.

Statistics South Africa (Stats SA) will not be liable for any damages or losses, except to the extent that such losses or damages are attributable to a breach by Stats SA of its obligations in terms of an existing agreement or to the negligence or willful act or omissions of the Stats SA, its servants or agents, arising out of the supply of data and or digital products in terms of that agreement. The user indemnifies Stats SA against any claims of whatsoever nature (including legal costs) by third parties arising from the reformatting, restructuring, reprocessing and/or addition of the data, by the user.

The data were gathered in October 2001. Since then, there have been demographic changes in South Africa associated, inter alia, with internal and external migration, and population growth. This means that population profiles may have changed at differing geographic levels. Stats SA is not responsible for any damages or losses, arising directly or consequently, which might result from the application or use of these data.

12. CONTACT DETAILS

Please do not hesitate to contact Stats SA User Information Services for additional information or queries:

Tel: +27 (12) 310-8600
Fax: +27 (12) 310-8500
E-mail: info@statssa.gov.za
Stats SA website: www.statssa.gov.za