

Census 2001: Metadata

INTRODUCTION

Overview of Census 2001	2
Census questionnaires	2
Instructions for interviewers	2
Adjusting for undercount	2
Confidentiality	3
Imputation	3
Statistics Council Census sub-committee comment	5
Copyright notice and disclaimer	5
Definitions	6
Abbreviations	6
Contact details	6

INTRODUCTION

Overview of Census 2001

Every person, young or old, present in South Africa on Census Night, 9-10 October 2001, should have been enumerated in Census 2001. The enumeration primarily took place over the period 10 October to 30 October. However, in some situations it was necessary to continue enumeration through to November to ensure that as many people as possible were included. Most respondents were interviewed, although in certain circumstances respondents were asked or allowed to complete the questionnaire themselves. Self-enumeration occurred in fewer than five per cent of cases.

Census questionnaires

Three different questionnaires were used. The differences between these are noted in the documentation on each census variable and can be seen in the electronic versions of the questionnaires provided. The three questionnaires were as follows:

1. QUESTIONNAIRE A (FOR HOUSEHOLDS)

Every household living in a housing unit, and every individual or household living in a workers' hostel, student residence, residential hotel or home for the independent aged, was enumerated on a household questionnaire. This obtained information about the household and each person present in the household on Census night, as well as about services available to the household.

Questionnaire A was administered in hostels and the other types of collective living quarters listed above, in order to capture any households resident in such places. In addition, staff and their households living on the premises of an institution (as opposed to occupants of the institution) completed Questionnaire A.

The metadata will use the term 'households' when referring to Questionnaire A.

2. QUESTIONNAIRE B (FOR INSTITUTIONS)

Questionnaire B was used to collect information on individuals who spent census night in institutions and tourist hotels. By institutions in this sense is meant hospitals, childcare institutions, boarding school hostels, homes for the disabled, initiation schools, convents, defence force barracks, prisons, community and church halls, and refugee camps. Homeless people were also enumerated with Questionnaire B. This questionnaire contains most of the questions contained in the first part of Questionnaire A, with the exception of question about family relationships. It has no questions about services. One questionnaire was completed for each individual in an institution and tourist hotel and for each homeless person found on the street.

The metadata will use the term 'institutions' when referring to Questionnaire B.

3. QUESTIONNAIRE C (FOR INSTITUTIONS)

Questionnaire C asked questions on housing and services for each institution and tourist hotel, and was completed with the cooperation of the manager or responsible person. In addition, all persons enumerated using Questionnaire B were listed on this questionnaire.

The metadata will use the term 'institutions' when referring to Questionnaire C.

Instructions for interviewers

All interviewers were given an Enumerator's Manual that included a section with instructions for each question on the questionnaires. Some extracts from these instructions have been included in the documentation as an indication of the theoretical basis of the questions. However, it is important to note that not every interviewer would have been completely familiar with all the instructions and therefore they do not necessarily reflect how the questions were asked by the interviewer. In addition, these instructions were not provided to people who completed the questionnaire themselves. In most cases the latter would have been required to give their own interpretation to questions. A single page of guidelines for self-enumeration was produced but was not always available.

Adjusting for undercount

In every census, there are bound to be some people, households, or even entire EAs, that are missed, or some people who are counted twice. During November 2001, a post-enumeration survey (PES) was undertaken to determine the degree of undercount or overcount in Census 2001. For those who are interested in the details, a separate publication describing the methodology of the PES will also be available. See the publication *Calculating the undercount in Census 2001*, Statistics South Africa (ref. no. 03-02-17 (2001)). This publication will be available upon request from Statistics South Africa as well as being on the website.

The numbers and percentages relating to households and hostels in all Census 2001 products are adjusted according to the PES findings through the application of weights. The exception is the sample database product, which provides raw and weighted data for a small sample of questionnaires. Data relating to other collective living quarters are not weighted, as the PES did not cover these places.

Confidentiality

All cells in tables referring to person information with values less than 4 are made confidential by replacing the number in the cell with a random number between 0 and 3. All post-aggregation tabular derivations are calculated from the randomly rounded values and are not rounded further. With this method, there is no difference between totals as they appear in the table and the arithmetic sum of the total's components. However, totals suffer from accumulated rounding errors and may not match values given for the same total in other tables.

Imputation

Imputation was used in Census 2001 to allocate values for unavailable, unknown, incorrect or inconsistent responses. The editing system used a combination of both 'logical' imputation techniques and 'hot decks' (dynamic imputation) when inconsistencies were found in the data. 'Undetermined' values were allowed for only a few variables, such as industry and occupation. The system tried to make the minimum number of imputations in order to remove errors and make the data consistent.

Logical imputations, in which a consistent value is calculated or deduced from other information relating to the individual or household, are usually preferred over hot deck imputations. Generally, the editing system attempts to resolve inconsistencies first by looking at other characteristics of the household or individual (for example, a married person with invalid sex would be assigned to the opposite sex of his or her spouse). If this is unsuccessful, then a consistent value is imputed from a hot deck.

More information on imputation can be obtained in the United Nations publication *Handbook on Population and Housing Census Editing*, reference: ST/ESA/STAT/SER.F/82. In addition, the complete set of editing specifications will be available on Stats SA's website www.statssa.gov.za (Ref: 03-02-43 (2001)).

The following table shows the imputation rate for each variable:

Variable	Imputation rate %	Universe
Age	23,72	All persons
Year of birth	2,34	
Sex	1,25	All persons
Relationship	5,11	All persons living in housing units and non-institutional collective living quarters
Marital status	5,50	All persons
Population group	1,42	All persons
Language	1,02	All persons
Religion	2,97	All persons
Born in South Africa	2,48	All persons
Place of birth	3,84	All persons born in South Africa
Country of birth	3,75	All persons not born in South Africa
Citizenship	4,44	All persons
Usually live	7,28	All persons
Province of usual residence	47,30	All persons that are not usual residents
Main place of usual residence	31,48	All persons that are not usual residents
Five years ago	5,71	All persons
Province of previous residence	19,02	All persons that have moved since the time of census 1996
Mother alive	1,90	All persons
Father alive	3,41	All persons
Present school attendance	7,51	All persons
Type of institution	9,58	All persons attending an educational institution
Level of education	8,60	All persons aged 5 years and over
Field of education	19,85	All persons with level of education higher than matric
Any work 7 days before 10 October	6,90	All persons aged 10 years and older

Variable	Imputation rate %	Universe
Reason why not working	4,73	All persons aged 10 years and older who did not have work
Active steps	17,02	All persons aged 10 years and older who did not have work
Availability	25,91	All persons aged 10 years and older who did not have work
Work status	11,03	All persons aged 10 years and older who worked
Economic sector	9,62	All persons aged 10 years and older who worked
Occupation	6,60	All persons aged 10 years and older who worked
Hours worked	8,13	All persons aged 10 years and older who worked
Place of work	12,13	All persons aged 10 years and older who worked
Province of work	38,07	All persons aged 10 years and older who did not work in same sub-place
Main place of work	4,72	All persons aged 10 years and older who did not work in same sub-place
Total births	17,43	All women aged 12-50 years
Total births (boys)	22,87	All women aged 12-50 years
Total births (girls)	23,04	All women aged 12-50 years
Boys still living	34,99	All women aged 12-50 years
Girls still living	35,46	All women aged 12-50 years
Sex of last birth	22,03	All women aged 12-50 years
Last child still alive	19,34	All women aged 12-50 years
Travel to school or place of work	15,94	All persons
Income	15,75	All persons
Type of living quarters	3,10	All households
Type of housing unit	1,36	All households in housing units
More than one dwelling	1,91	All households in housing units
Rooms	1,83	All households in housing units
Sharing 1 room	8,99	All households in housing units with one room
Tenure status	2,54	All households in housing units
Piped water	2,17	All households and institutions
Source of water	0,87	All households and institutions
Toilet facility	0,76	All households and institutions
Energy source for cooking	1,38	All households and institutions
Energy source for heating	3,23	All households and institutions
Energy source for lighting	0,92	All households and institutions
Household goods – radio	0,88	All households and institutions
Household goods – television	0,86	All households and institutions
Household goods – computer	1,93	All households and institutions
Household goods – refrigerator	0,89	All households and institutions
Household goods – telephone	1,38	All households and institutions
Household goods – cell-phone	1,52	All households and institutions
Access to telephone	5,08	All households and institutions with no telephone or cell-phone
Refuse or rubbish disposal	2,39	All households and institutions
Anybody died	3,58	All households and institutions
Sex of deceased	1,91	All deceased
Age of the deceased	7,75	All deceased
Cause of death	5,20	All deceased
Deceased pregnant	18,15	All deceased women aged 12-50 years

Statistics Council Census sub-committee comment

“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*.”

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 wilful 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 the data gathered as part of the 2001 population census.

Definitions

See Census 2001 *General Manual of Concepts and Definitions* (ref. no. 03-02-03).

Abbreviations

A number of abbreviations have been used in this documentation. They are:

Cat A	Category A municipalities (Metropolitan Areas)
Cat B	Category B municipalities (Local Municipalities and District Management Areas)
Cat C	Category C municipalities (District Councils)
CBLC	Cross-boundary Local Council
CBDMA	Cross-boundary District Management Area
DC	District Council
DMA	District Management Area
EA	Enumeration Area
EC	Eastern Cape
Excl.	Excluding
FS	Free State
GIS	Geographic Information System
Govt	Government
GP	Gauteng
ID	Identification
KN	KwaZulu-Natal
Metro	Metropolitan area
Mnfctr.	Manufacturing
MP	Mpumalanga
Munic.	Municipality
N/A	Not applicable (that is, no response was required to the particular question from these households or persons)
NC	Northern Cape
NEC	Not elsewhere classified (detailed information provided by respondent, but classification does not include a detailed code for the response)
NFD	Not further defined (insufficient information provided by respondent to enable a more detailed code to be allocated)
No.	Number
NP	Limpopo (former Northern Province)
NW	North West
PES	Post-enumeration Survey
WC	Western Cape

Terms requiring further explanation:

Pre-coded	Questions where a list of responses was offered for the respondent to choose from.
Write in	Questions where the respondent/interviewer had to provide an exact response. Often a precise code was allocated at a later stage.

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