



Statistics South Africa

CENSUS 2001

How the count was done



Statistics South Africa
Private Bag X44
Pretoria 0001
South Africa

170 Andries Street
Pretoria 0002

User information service: (012) 310 8600
Fax: (012) 310 8500
Main switchboard: (012) 310 8911
Fax: (012) 322 3374

Website: www.statssa.gov.za
E-mail: info@statssa.gov.za

CENSUS 2001

How the count was done

Statistics South Africa
2003

Report no. 03-02-02 (2001)

Pali Lehohla
Statistician-General

Published by Statistics South Africa, Private Bag X44, Pretoria 0001

© **Statistics South Africa, 2003**

Users may apply or process this data, provided Statistics South Africa (Stats SA) is acknowledged as the original source of the data; that it is specified that the application and/or analysis is the result of the user's independent processing of the data; and that neither the basic data nor any reprocessed version or application thereof may be sold or offered for sale in any form whatsoever without prior permission from Stats SA.

ISBN 0-621-34294-7

Stats SA Library Cataloguing-in-Publication (CIP) Data

Census 2001: How the count was done / Statistics South Africa. Pretoria: Statistics South Africa, 2003

Previous title: The people of South Africa: Population census, 1996: The count and how it was done

55p. [Report No. 03-02-02 (2001)]

ISBN 0-621-34294-7

1. Census
2. Census undercounts
3. Lists
4. Census – planning
5. Census – methodology

I. Statistics South Africa

II. Series

(LSCH 16)

A complete set of Stats SA publications is available at the Stats SA library and the following libraries:

National Library of South Africa, Pretoria Division
National Library of South Africa, Cape Town Division
Library of Parliament, Cape Town
Bloemfontein Public Library
Natal Society Library, Pietermaritzburg
Johannesburg Public Library
Eastern Cape Library Services, King William's Town
Central Regional Library, Polokwane
Central Reference Library, Nelspruit
Central Reference Collection, Kimberley
Central Reference Library, Mmabatho

This report is available on the Stats SA website: www.statssa.gov.za

Copies are available from: Printing and Distribution, Statistics South Africa

Tel: (012) 310 8251

Fax: (012) 322 3374

E-mail: distribution@statssa.gov.za

Contents

List of tables	v
List of figures	v
Definitions	vi
Overview and summary of results	
Census 2001: counting the people	1
Population of South Africa in total and by sex	2
Population by province	2
Comparison by province and population group	3
Summary	4
Chapter 1: Planning Census 2001	
Initiating the process	5
The census programme	6
Subcontracts and public–private partnerships	8
Chapter 2: Pre-enumeration	
Demarcation and map production	9
The census questionnaires	10
Preparation for the post-enumeration survey	11
Initial training plans	12
Testing procedures in the field	12
Problems identified	13
Changes introduced	13
Listing	14
Establishing structures within each province	15
Selection and appointment of temporary staff	16
Plans for payment	17
Preparation for data processing	17
Distribution of questionnaires to regional offices	18
Public outreach and publicity	18
Amended training method	19
Chapter 3: Enumeration	
Overview	21
Co-ordination and quality control	22
Steps taken to increase coverage	23
Payment issues	24
Handling the sealed boxes	24

Chapter 4: Data processing	
Introduction	25
Data capture	25
Coding	27
Post-capture processing	28
Summary	28
Chapter 5: The post-enumeration survey	
Methodology	31
Preparation	32
PES enumeration	33
The matching process	33
The adjustment process	35
Chapter 6: Key results	
The population of South Africa	37
The households of South Africa	44
Summary of adjustments	44
Appendix A: Members of the Statistics Council	46
Appendix B: PES adjustment	47
Appendix C: Census products	48

List of tables

Table A: The South African population in October 2001 and October 1996	2
Table B: Estimates of the South African population in October 2001 and October 1996, by province	3
Table C: Estimates of the South African population in October 2001 and October 1996, by population group	4
Table 2.1: Provincial organisational structure for Census 2001	16
Table 5.1: Percentage undercount for persons and households in each province	35
Table 6.1: The population of South Africa by province and sex	37
Table 6.2: The population of South Africa by population group within provinces	38
Table 6.3: The population of South Africa by province within each population group	39
Table 6.4: Age distribution of South Africans in five-year intervals	40
Table 6.5: Households in South Africa by type of dwelling and population group of the household head	44

List of figures

Figure 6.1: Distribution of the total population by age group and sex	41
Figure 6.2: Distribution of the black African population by age group and sex	42
Figure 6.3: Distribution of the coloured population by age group and sex	42
Figure 6.4: Distribution of the Indian or Asian population by age group and sex	43
Figure 6.5: Distribution of the white population by age group and sex	43

Definitions

A publication of concepts and definitions used in Census 2001 will be available electronically. The main terms used in this publication are defined below.

- An *enumeration area (EA)* is the smallest geographical unit (piece of land) into which the country is divided for census or survey enumeration, of a size able to be enumerated by one census fieldworker (*enumerator*) in the allocated period. EAs typically contain between 100 and 250 households.
- An *enumerator* is the person who visits each household and other individuals in a specific EA and administers the questionnaires or arranges for self-enumeration; a census fieldworker.
- A *household* is a group of persons who live together, and provide themselves jointly with food and/or other essentials for living, or a single person who lives alone. (The 'four-night-a-week' criterion for household membership does not apply, as this was a *de facto* census, that is, people were counted where they were staying on census night.)

Note: In this census, household information was collected from households in housing units and also from households in the following living quarters: workers' hostels, student hostels, homes for the independent aged and residential hotels. However in this volume, results are given for households in housing units only.

- *Housing unit:* A unit of accommodation for a household, which may consist of one structure, more than one structure, or part of a structure. Examples of each are a house, a group of rondavels, and a flat.
- *Collective (or communal) living quarters:* living quarters where certain facilities are shared by groups of individuals or households. They include hostels, hotels and institutions.
- *An institution:*
 - (1) *enumeration* – a communal place of residence for people with a common characteristic, such as a hospital, school hostel, prison, defence force barracks or convent. Residential hotels, workers' hostels, students' residences and homes for the independent aged are not counted as institutions in this sense. (This group is sometimes referred to as 'non-institutional collective living quarters'.)
 - (2) *demarcation* – one of ten geographical EA types. All collective living quarters other than hostels were classified as institutions for demarcation purposes. This use of the word institution is therefore broader than the first meaning.

Overview and summary of results

*Sibale sonke, hihlayeni hinkwerhu, re bale kaofele, re bale kamoka,
re bale ro rotlhe, ri vhaleni rothe, tel ons by, count us in.*

Census 2001: counting the people

During October 2001 the second census in a democratic South Africa took place. Approximately 100 000 temporary employees of Statistics South Africa visited the various metros, cities, towns, townships, informal settlements, villages, farms and deep rural areas throughout the country. Their task was to record the details of the people who were present in the country on the night of 9–10 October 2001. People living in households across the country, as well as those in hostels, hotels, hospitals and all other types of communal living quarters, and even the homeless, were all visited.

In preparation for the count, the entire country had been divided into approximately 80 000 'parcels' of land called enumeration areas (EAs), each containing an average of 150 households, or in the case of communal living quarters, an equivalent workload. One enumerator was allocated to each EA to visit all the households and individuals in the EA and complete a questionnaire or leave a questionnaire to be filled in.

Immediately after the count, fieldworkers in the provinces, working independently of the census, re-visited a representative sample of 600 enumeration areas, identified their boundaries, listed afresh the dwellings, and visited the households again to complete another questionnaire. Hostels were also revisited but no other communal living quarters. The questionnaire contained a subset of the census questions and also sought to ascertain the whereabouts of each household member or hostel-dweller on census night.

This exercise, called a 'post-enumeration survey' or PES, is conducted in many countries following a census. The results from the post-enumeration survey are compared with those from the census by matching the corresponding questionnaires, and where there is uncertainty, revisiting the specific household, the particular EA, or even the surrounding EAs. This careful process reveals the extent of households and people missed or counted more than once in the main census. Necessary adjustments to the figures from the main census are then calculated.

The results that follow are based on the actual count of Census 2001, but are adjusted for undercount according to the findings of the PES.

Population by sex

The final estimates of the size of the population of South Africa as at October 2001, after adjustment for undercount, are shown in Table A below. By way of comparison, the final estimates for October 1996 are also shown.

Table A: The South African population in October 2001 and October 1996

	October 2001		October 1996	
	N	%	N	%
Males	21 434 040	47,8	19 520 887	48,1
Females	23 385 737	52,2	21 062 685	51,9
Total*	44 819 778	100,0	40 583 573	100,0

* All PES-adjusted numbers are rounded to whole numbers. The totals displayed may therefore differ slightly from the sum of the separate numbers.

It is seen from Table A that the population of South Africa in 2001, according to the final estimate after adjustment for undercount, was 44,8 million, as against 40,6 million in 1996.

It is also seen that the proportions of men and women in the population in 2001 were 47,8% and 52,2% respectively, according to the final estimates.

Population by province

Table B below shows the population estimates for October 2001 and October 1996 by province. It indicates that KwaZulu-Natal had the largest population (9,4 million people) while the Northern Cape had the smallest (0,8 million people).

Table B: Estimates of the South African population in October 2001 and October 1996, by province

Province	October 2001		October 1996	
	N	%	N	%
Eastern Cape	6 436 763	14,4	6 302 525	15,5
Free State	2 706 775	6,0	2 633 504	6,5
Gauteng	8 837 178	19,7	7 348 423	18,1
KwaZulu-Natal	9 426 017	21,0	8 417 021	20,7
Limpopo	5 273 642	11,8	4 929 368	12,1
Mpumalanga	3 122 990	7,0	2 800 711	6,9
Northern Cape	822 727	1,8	840 321	2,1
North West	3 669 349	8,2	3 354 825	8,3
Western Cape	4 524 335	10,1	3 956 875	9,7
South Africa*	44 819 778	100,0	40 583 573	100,0

* All PES-adjusted numbers are rounded to whole numbers. The totals displayed may therefore differ slightly from the sum of the separate numbers. Similarly, the percentages are rounded to the first decimal place, and therefore may not always add up to exactly 100.

The increase in the proportion of the population living in Gauteng and the Western Cape, and the decrease in the proportion living in the Eastern Cape, Limpopo, the Free State and the Northern Cape, is suggestive of migration from the more rural to the more urban provinces.

Population by population group

Table C overleaf shows the breakdown of the population into the four main population groups according to the final estimates of Census 2001, alongside a comparison with Census '96.

The proportion of African people increased from 76,7% in October 1996 to 79,0% in October 2001, while the proportion of white people decreased from 10,9% in October 1996 to 9,6% in October 2001. The proportions of coloured people (8,9% in October 1996 and 8,9% in October 2001) and Indian or Asian people (2,6% in October 1996 and 2,5% in October 2001) remained approximately the same. Note, however, that in 1996 there was a small group classified as 'other or unspecified'.

Table C: Estimates of the South African population in October 2001 and October 1996, by population group

Population group	October 2001		October 1996	
	N*	%**	N*	%**
Black African	35 416 166	79,0	31 127 631	76,7
Coloured	3 994 505	8,9	3 600 446	8,9
Indian or Asian	1 115 467	2,5	1 045 596	2,6
White	4 293 640	9,6	4 434 697	10,9
Other or unspecified*			375 204	0,9
Total**	44 819 778	100,0	40 583 573	100,0

* In 2001, logical and dynamic imputation was used to reclassify the few people who did not indicate their population group or who described themselves as something other than one of the four options given.

** All PES-adjusted numbers are rounded to whole numbers. The totals displayed may therefore differ slightly from the sum of the separate numbers.

Summary

During the count of Census 2001, as adjusted by the PES, Stats SA found that there were 44,8 million people living in South Africa on the night of 9/10 October 2001. Of these:

- 21,4 million (47,8%) were male, and 23,4 million (52,2) female.
- 35,4 million (79,0%) were African, 4,0 million (8,9%) coloured, 1,1 million (2,5%) Indian/Asian, and 4,3 million (9,6%) white.

The province with the largest number of people (9,4 million or 21,0%) was KwaZulu-Natal, followed by Gauteng (8,8 million or 19,7%), the Eastern Cape (6,4 million or 14,4%), Limpopo (5,3 million or 11,8%), the Western Cape (4,5 million or 10,1%), North West (3,7 million or 8,2%), Mpumalanga (3,1 million or 7,0%), and the Free State (2,7 million or 6,0%), while the province with the smallest number of people was the Northern Cape (0,8 million or 1,8%).

The population of South Africa had increased by 4,2 million people between October 1996 and October 2001.

Chapter 1: Planning

Initiating the process

Planning for Census 2001 began in earnest in November 1999, after the necessary funding had been obtained through the parliamentary budget process. This funding was relatively late in coming, which left little time for the careful planning that is ideally required. For example, in Canada, which also conducts a census every five years, work begins seven years in advance, before the previous census has even taken place.

In view of the short time-frame, alternatives to conducting a full-scale census were investigated. For example, Stats SA investigated the viability of conducting a large-scale household survey, in which every sixth household in the country would be visited. But the census advisory committee, which included members of the Statistics Council, rejected this alternative in favour of a full-scale census.

Firstly, it was felt that the country needed detailed, updated information to measure progress against the baseline data of Census '96. Census '96 had been the first fully comprehensive South African census since 1970, as in the intervening years the 'independent states' and the homelands had not been included.

Secondly, the census advisory committee felt that it would be just as difficult logistically, and just as expensive, to conduct a large-scale survey. For example, in many parts of the country people do not have easily identifiable physical addresses, and in other places, roads are poor or absent, making access difficult. A decision was therefore taken to proceed with the census.

The census advisory committee

In order to ensure that a wide variety of stakeholders and users participated in the census, an advisory committee was set up. It consisted of diverse interest groups, including academics, representatives of civil society, other government departments, parastatal organisations, non-governmental organisations and the business community. The committee met at intervals throughout the preparatory and enumeration phases of Census 2001. The members of the committee gave advice on aspects of the census, including questionnaire design.

The Statistics Council

A census sub-committee of the Statistics Council was created. This sub-committee gave advice on questionnaire design and the enumeration process. It also reviewed the overall census methodology and evaluated the findings after the exercise was completed. A list of the members of the council and the census sub-committee can be found in Appendix A.

International assistance

Extensive assistance with strategy, census mapping, questionnaire design, fieldwork, manuals and data processing was given by census specialists from the United States Census Bureau, funded by the United States Aid for International Development (USAID), and from Stats Sweden, funded by the Swedish Institute for Development Aid (SIDA). Consultants from Kenya, Tanzania and the UK also assisted with various aspects of data processing.

The census programme

Census 2001 was originally conceived as a programme consisting of various distinct sub-projects, which required a project management approach to identify specific tasks, timelines and risks. This approach differed from the one taken in Census '96, where the various tasks were dealt with sequentially, and problems were solved on a more ad hoc basis.

In common with Census '96, Census 2001 was conceptualised as consisting of the following four phases: pre-enumeration, enumeration, data processing, and analysis and dissemination. However, the links within and between these phases were more clearly defined in 2001 than in 1996.

Contracts with the private sector were considered for the first time in 2001. In 1996, Stats SA handled all operations, including recruiting and managing its own staff and developing all computer systems and programmes.

The original programme of Census 2001 was broken down into a first and second level hierarchy of sub-projects in the following way:

First level: project office

At this level, a project office was established as the main co-ordinating structure to manage the project. Each sub-project was required to report back to the project office on plans, implementation procedures, deadlines and achievements, according to specified guidelines. This method of co-ordination proved to be insufficient, since tasks and timelines were emphasised at the expense of the content of each sub-project and the inter-relationships between them.

Second level: sub-projects

Operational sub-projects

Eight operational sub-projects were identified, each reflecting an aspect of the line function of the entire census process, and each requiring detailed planning and implementation. These sub-projects were:

- Questionnaire design
- Census mapping (demarcation)
- Geographical information system (GIS) updating and maintenance
- Design and completion of the pilot census
- Enumeration
- Post-enumeration survey (PES) design and execution
- Data processing
- Data tabulation and product planning and implementation.

Support sub-projects

In order for these line-function projects to operate, a series of administrative and support sub-projects were also instituted. These were:

- Financial management and monitoring
- Information technology
- Human resources
- Provisioning
- Logistics
- Provincial management
- Publicity.

Subcontracts and public–private partnerships

At the same time, it was decided to outsource two essential functions or sub-projects: payment of the temporary workers, and the technical and management aspects of data processing. This decision was taken in view of the major difficulties experienced with these two components in Census '96. A private communications company was contracted to share the publicity task, whilst contractors were also used to assist with demarcation in various parts of the country.

Later, a private management consultancy was introduced to assist in integrating the various components of the census structure. This consultancy was also responsible for developing the computer-based management system called the Census Administration System (CAS), and for the distribution of boxes to the provinces and regional offices.

Even though there were some successes, particularly in relation to the management consultancy that assisted with the integration, several new and unexpected problems were experienced in working with the private sector, some of which are highlighted below.

Chapter 2: Pre-enumeration

Demarcation and map production

In the pre-enumeration phase, Statistics South Africa had to demarcate the whole country into small geographic areas, called enumeration areas (EAs). The underlying principle is that all parts of an EA should be within comfortable reach of an enumerator, so that enumeration of all households can be completed within the allocated number of days.

To complete this undertaking, three distinct operations were performed: (1) creation of a spatial information database, (2) demarcation of EAs, and (3) printing of maps for enumerator summary books.

For the first time, Geographic Information System (GIS) technology was utilised to demarcate EAs and for map production instead of the traditional methods of using analogue and sketch maps. A comprehensive digital spatial information database was created from several data sets acquired from government departments and private sector companies. The data sets included topographic maps, cadastral data, administrative boundaries, aerial photography, satellite imagery and videography. In areas where digital data was not sufficient, Global Positioning Systems (GPS) were used to collection information in the field, which was then incorporated into the spatial information database. These data sets were all integrated into one common spatial frame. Special attention was focused on areas that were difficult to demarcate, particularly tribal areas and urban informal areas. This large spatial information database, approximately 1,8 terabytes, provided demarcators with the most complete and current spatial information collection to create EAs.

A geographical hierarchy structure was created to link EAs to administrative boundaries. The EA is the smallest geographic unit and can be aggregated up to municipal, district and provincial level. This geographical hierarchy structure is fundamental in demarcating EAs and in disseminating census information.

For EAs to be functional, demarcation must adhere to certain specifications regarding administrative and social boundaries, shape, size, population density, and the mobility of enumerators at the time of the census. The availability of spatial data played a dominant role in the evaluating of the old 1996 EA boundaries and areas, as well as in identifying the changes that had to be made. For the first time, on-screen demarcation was possible, saving a great deal of time and effort.

After delineating the boundaries of the EAs in such a way that they were recognisable on the ground, each EA was classified according to its location as one of four types: formal urban area, informal urban area, rural area (commercial farms) or tribal (traditional) area. Depending on the dominant land use, EAs were sub-typed as being residential, farm, small holding, recreation/park/state land, institution, hostel or vacant. Attribute data were then collected according to the sub-type of the EA. Information such as the place name, the type of settlement, and the type and name of institution was captured. Most of this information was not digitally available and was collected by demarcation field teams.

Each EA received a unique code number, linking it to its geographical entities. All subsequent census processes used this number as an identifier. Supervisor Units (SUs) were created using the EA sub-type as a basis. Approximately five adjacent EAs of similar type were grouped together as units to be managed by a supervisor during enumeration.

Map production is an important census activity. Maps were produced for every EA (81 000 A3 maps), supervisor area (16 000 A3 maps), and regional and provincial office (100 A0 maps). Each enumeration summary book contained at least one EA map and an orientation map. The maps were laminated for protection against damage and bound into the summary books. Regional and provincial maps were created to assist with planning and controlling census operations. Bar codes were added to the map layout for control purposes, to ensure that the correct EA and supervisor maps were attached. Thick paper stock was used to allow for double-sided colour printing.

The census questionnaires

The pre-enumeration phase also involved the development and testing of the questionnaire. The Census '96 questionnaire was used as a basis. Each provincial office held a series of meetings with stakeholders to decide on the content. While some questions were similar to those asked in Census '96, others, for example those relating to the labour market, were substantially different. As in Census '96, the need for detailed information at local level was one of the main factors contributing to the length of the questionnaire. Questions and layout were tested with behind-the-glass interviews towards the end of 2000.

Three different census questionnaires were developed – one for households (the A questionnaire), one for individuals in institutions (the B questionnaire), and one for the institutions themselves (the C questionnaire). The A questionnaire was also used in workers' hostels, student hostels, residential hotels and homes for the independent aged, whilst the B and C questionnaires were also used in tourist hotels and for the homeless.

The household questionnaire had space to collect information on ten people, whereas the B questionnaire had space for only one respondent. The personal questions were identical in the two questionnaires, with the exception of relationship questions, which were not relevant for individuals. The A questionnaire also contained questions on housing, services and household goods. These questions were not in the B questionnaire, but were asked in the C questionnaire in relation to the institution as a whole. The C questionnaire also provided for a list of all the residents on census night, to be supplied by the management of the institution, against which completed B questionnaires could be checked.

Due to the tight deadlines of the census project, there was an inherent risk that the local company subcontracted to print the questionnaires would be unable to produce sufficient quantity of an acceptable quality in time for distribution before the start of the count. An American firm was therefore subcontracted to produce about half of the questionnaires needed. These were flown into the country by jumbo jet direct to a South African air force base in Pretoria, where Stats SA had use of a warehouse.

Preparation for the post-enumeration survey

The methodology of the post-enumeration survey was also determined during this phase. Firstly, steps were taken to overcome the uncertainties that were experienced regarding the PES in the previous census. In 1996, Stats SA could not be sure whether 22% of the households visited during the PES had been visited during the actual count, even after an intensive matching process. There was also uncertainty as to whether certain individuals within households had been counted or not. A statistical model was therefore developed to estimate the probability of having counted each of the households and individuals about which there was uncertainty.

For the 2001 PES, Stats SA decided to conduct reconciliation visits for all cases that could not be matched when the census and PES data were compared. This was to include visiting adjacent EAs when necessary. Furthermore, extra questions, including

questions on in-movers and out-movers after the date of the count (the night of 9–10 October), were added to the 2001 PES questionnaire.

Another innovation was the use of barcode stickers on the census questionnaire to be left with the household and then transferred to the PES questionnaire, to facilitate matching. This process is described below.

Initial training plans

Training the temporary staff hired to carry out the census in the field is an enormous logistical task. The exercise inevitably involves some form of cascade training, whereby people are trained in the processes and methodology, and in turn train further groups of people, and so on down the line. This method is widely used in other countries, usually supported by training material to be used verbatim.

Stats SA hired eighteen national trainers who would be responsible for training the first level of staff in the provinces. These trainers were also involved in preparing training manuals from the reference manuals. Too much emphasis was put on training methods and not enough on census methodology, partly because the trainers did not fully understand the processes involved, and partly because they were not always kept informed of methodological decisions. Moreover the verbatim method was rejected as being inappropriate in the new South African educational context. For these reasons, as well as extreme time pressure, the manuals prepared for the pilot census were inadequate.

Testing procedures in the field

The census test

In August 2001 a census test was carried out primarily to test logistics and the PES questionnaire. It examined in particular the feasibility of using stickers to make the matching process between census and PES questionnaires easier. Each census questionnaire had two stickers with a unique barcode on a detachable strip. One sticker was placed on the door of the dwelling, while another was given to the respondent for safekeeping. If a particular dwelling was visited during the PES, the fieldworker attached the loose sticker to the PES questionnaire, or if necessary copied the barcode from the sticker attached to the door. The test showed that the use of stickers indeed assisted in the matching process.

The pilot census

The main pilot for the census was conducted in February and March 2001. All aspects of the enumeration phase were tested and it was viewed as a dress rehearsal for the main count. It was also used to develop the tools for data capture. The pilot confirmed the suspected deficiencies in integration and resulted in several major revisions of both processes and management methods.

Problems identified

The task-based project management approach used initially was shown by the pilot to hinder integration between census processes. The main difficulties identified were as follows:

- As the focus was mainly on timelines and task completion, plans did not sufficiently link all components of the census in a coherent management approach. This resulted in the duplication and overlapping of tasks, large gaps in planning, and in some cases, related tasks being tackled in non-compatible ways. For example, the planning of logistics and the planning of enumeration were not integrated. Communication between members of different teams was also inadequate.
- Enumerators' understanding of geography was generally inadequate. They struggled to read the maps and aerial photographs, were often unable to identify EA boundaries on the ground, and found the task of listing households in each EA difficult. There were also differing interpretations of the questionnaires. It was clearly necessary to improve plans for training enumerators.
- The original internally-developed administration system was inadequate. It was not able to capture all the necessary information and was not used correctly.

Changes introduced

After the pilot census, a new method of management was introduced to encourage greater integration across sub-projects. The structure did not alter significantly, but daily 'Nerve Centre' meetings were established, attended by the top census management and the various managers of the sub-projects as necessary. The aim of these meetings was to identify problems and take immediate action, having examined in particular how proposed methods or timing would impact on related processes. In addition, a Content, Development and Integration committee was formed, which later became the Integration and Implementation committee.

A private-sector consultancy attended the daily Nerve Centre meetings and assisted in the overall management of this new process. The Nerve Centre members took all strategic and tactical decisions, and these were passed on to the provincial heads for implementation by means of regular video-conferencing sessions.

The main changes were as follows:

- It was decided that it was not possible to expect enumerators to list EAs prior to enumeration. Listing became a separate sub-project.
- A new way of training and cascading the training downwards was initiated, using 'narrowcasting' for the training of trainers and video equipment for the training of regional staff, supervisors and enumerators. Overseas consultants and seconded staff assisted with developing manuals.
- The census information and management system, called the *Census administration system* (CAS), was developed by the private-sector consultancy, who had previously successfully teamed up with the Independent Electoral Commission for this type of countrywide work.
- Barcodes were used to control the flow of questionnaire boxes to the regional offices, back to the data processing centre after enumeration, and within the data processing centre itself.

Listing

Listing takes place within each mapped EA. It involves making a list of all dwellings within the EA in the summary book provided, and indicating a suitable route to follow through the EA to ensure that all dwellings are visited.

In a formal urban area, dwellings are usually associated with a specific street address. In an informal settlement, however, very few dwellings have specific addresses. This is also the case in many rural areas, where dwellings may have to be identified by farm names, householders' names, or a description of the structure or its relationship to other structures. Listings by name are often completed with the help of the local headman or tribal authority. In a country where perhaps half of all households do not have a physical address, the listing of dwellings is not easy to achieve.

The original plan was that the enumerators or fieldwork supervisors would list the EAs immediately prior to enumeration. One reason for this approach is that in South Africa housing arrangements are constantly changing; new developments,

particularly informal settlements, spring up very rapidly, and lists of dwellings quickly become out of date.

However, as described above, the listing task was not well understood or carried out during the pilot census. Most of the temporary staff employed at enumerator level had no map-reading experience and found it difficult to relate the maps to the reality on the ground. It was therefore decided to employ separate specialised staff to list the dwellings in the EAs in advance of enumeration. It was also decided to reintroduce a written boundary description for each EA to make the enumerators' task easier, even though the maps and aerial photographs were vastly superior to the ones used in 1996. This was also to be done by the listers.

Listing therefore became a separate sub-project. Selected Head Office staff trained listers in different parts of the country on a rotation basis to ensure that the procedures were well understood. Eventually, when time grew short, televised narrowcast training was used. Not all areas in the country were completely listed by the time enumeration began, but there were several definite advantages to the revised plan, such as early identification of larger than anticipated EAs where extra enumerators would be required.

When their contracts had expired, the most capable listing staff were re-employed as fieldwork supervisors, since they had already gained experience of map reading and identifying EAs and their boundaries on the ground.

Establishing structures within each province

After the pilot census, the operational management of the census was put into effect through the establishment of 95 regional census offices throughout the country, each headed by a regional office manager. These managers reported to a provincial census manager who operated from within the permanent Stats SA provincial office and reported in turn to the Stats SA provincial manager. Each regional office had a logistics, finance, recruitment, IT and supervisory function.

Each regional office received computers for data capture and had CAS installed. They were linked to each other through a wide area network.

Stats SA then prepared to hire the following groups of temporary staff: approximately five staff members per regional office; about 16 000 staff for listing; and approximately 2000 fieldwork co-ordinators to run field stations, 16 000 supervisors, and 80 000 enumerators.

Table 2.1 provides a breakdown of the regional offices, supervisor units and enumeration areas in each province for Census 2001.

Table 2.1: Provincial organisational structure for Census 2001

Province	Number of regional offices	Number of supervisor units	Number of EAs
Eastern Cape	13	3 529	18 370
Free State	6	1 022	5 183
Gauteng	16	2 635	13 367
KwaZulu-Natal	16	2 480	12 752
Limpopo	6	322	1 661
Mpumalanga	7	1 129	5 813
Northern Cape	11	1 975	10 325
North West	7	1 188	6 215
Western Cape	13	1 379	7 101
Total	95	15 659	80 787

Selection and appointment of temporary staff

Recruitment was decentralised to regional office level. Advertising of census positions took place through print media and radio, through notices at municipal offices and at provincial and regional office launches. Enumerators were expected to have matric, literacy in English, map-reading ability, and knowledge of the language of the area. Completed application forms were screened using a predetermined scoring system. All information from the application forms was entered into the new Census Administration System (CAS), which produced a total score for each applicant and created a shortlist of qualifying applicants. Problems with the CAS meant that not all appointments were made on time, which in some areas led to a late start to enumeration.

Plans for payment

The following plans were originally made for the payment of temporary census staff:

- Regional office and fieldwork co-ordination staff would be paid through the normal government (Persal) pay system, as they generally had longer contracts than the other temporary staff.
- Payment of the listing staff, fieldwork supervisors and enumerators was subcontracted. They would be paid through the Post Office.
- Details of each appointment would be captured at regional level through the CAS.

Preparation for data processing

Development of the data-capturing system by the private contractor who won the tender started after the pilot census. For the first time, scanning was to be used to capture each page of the questionnaire as an image, linked to software to interpret the images.

Further preparation for data processing involved the development of:

- electronic coding dictionaries for coding the open-ended questions;
- tabulation plans to ensure that the data would be captured in ways appropriate for analysis; and
- software for post-capture data-transfer, editing and verification.

Logistical and human resources preparations included the following tasks:

- identifying and renting suitable premises and converting them for data processing;
- installing electrical cabling and air-conditioning for the scanning area and file server room;
- installing the scanners, file-servers, storage and memory systems, computer terminals and other hardware;
- installing the cabling to link all these electronic components;
- introducing a management system for the store;
- recruiting and training suitable people for all stages of data capture; and
- designing workflow and management methods.

Distribution of questionnaires to regional offices

- The printed questionnaires were all delivered to the warehouse in Pretoria. The questionnaires each had a unique barcode which was used throughout enumeration and data processing for tracking purposes.
- The questionnaires were distributed in enumerator boxes. Each box contained four packs of fifty household questionnaires, sealed in plastic, and one pack each of B and C questionnaires (for use in institutions), as well as materials required for enumeration such as pencils, bibs and caps. Each box also had a barcode, which was scanned into a central register before the box left the warehouse.
- The boxes were delivered to the regional offices where the barcodes were again scanned as a record of receipt. In this way, each regional office was responsible for specific questionnaires assigned to it.

Public outreach and publicity

A publicity campaign was launched in the mainstream media. The campaign initially focused on non-urban areas of the country through radio advertising and talk shows. The focus was extended to television and the print media in the final stages of the campaign. A communications company with associates in all nine provinces co-ordinated the campaign, which was overseen by the Stats SA census publicity sub-project.

The mainstream media campaign was complemented by a grassroots campaign which included provincial launches, visits to local authorities, and meetings with civic and traditional leaders.

Further publicity was given to the census by **Census at school**, a statistical literacy project conducted at all government-registered schools throughout the country during September. Children were encouraged to discuss the exercise in their households, thus raising awareness of the process of collecting information in a census.

Amended training method

The fieldwork co-ordinators, together with eighteen regional trainers appointed to oversee census training in each province, were first trained at central venues in each province by the trainers from Head Office, principally by means of narrowcasting onto television screens. Senior staff responsible for the census did the actual televised narrowcast training. The trainees watched the sessions live and were able to phone in to a help line to ask questions, which were answered immediately by the senior staff member. This ensured that the same methodology was given to all trainees around the country.

The fieldwork co-ordinators then undertook the training of supervisors and enumerators in their allocated areas, using videotapes of these live sessions as well as the significantly improved combined reference and training manuals. This overall method seemed to work better than using academics around the country, as was done in 1996. It also circumvented the multiple cascade system to a considerable degree, as supervisors did not have to train enumerators. Furthermore, removing listing from the enumerators' tasks gave them more time to familiarise themselves thoroughly with the questionnaires during the limited training period.

PLEASE SCROLL DOWN

Chapter 3: Enumeration

Overview

In the enumeration phase, fieldworkers called enumerators visited the households in the EAs throughout the country and ensured that a questionnaire was completed giving information on persons in the household. They also visited hostels, hotels and student residences, as well as institutions such as prisons, police cells, hospitals, homes for the disabled, and army barracks. The homeless were enumerated where they were found on the street.

The EA summary book gave a clear indication of the area in which each enumerator was required to work. As well as the maps, the book contained a list of all the dwellings in the EA, and their addresses if applicable, or else some other identifying description, and a route to follow through the EA. In the case of communal living quarters, each room, ward, cell, dormitory or section was listed. For each listed dwelling or equivalent, the enumerators were required to indicate in the book whether or not it was visited, and if so, the number of males and females counted in it. If the dwelling was not visited, a reason had to be given, for example, a vacant dwelling. Any extra dwellings found were to be added to the list and also enumerated.

Each enumerator was required to produce one or more completed questionnaires for each dwelling visited. Households were encouraged to be interviewed by the enumerator if possible. Alternatively, a respondent could complete the questionnaire for collection later, where circumstances allowed. Enumerators carried translations of the questions into the other ten official languages, to refer to where necessary. Some questionnaires were printed in the other ten languages, particularly Afrikaans and Zulu, but distributing them to the enumerators who needed them proved complicated.

Census 2001 was a *de facto* census, which means that people were enumerated where they stayed on census night, or, if they were not at a dwelling on census night and were not enumerated elsewhere, where they returned to the next day. (This applied to people at work, at places of entertainment, or at events such as night vigils.)

The supervisors were required to check the work of the enumerators, the fieldwork co-ordinators were required to check the work of the supervisors, and in turn the regional managers were required to check the work of their region.

Co-ordination and quality control

The supervisor's role was to ensure that the five or so enumerators under his or her control completed their work and submitted accurate completed questionnaires and summary books. (After the pilot census, the number of enumerators to be overseen by one supervisor was reduced from ten to five.) Among other tasks, the supervisor had to:

- hand out and receive back all materials, including blank and completed questionnaires;
- check each returned completed questionnaire for omissions or anomalies, and attempt to have them rectified in the field;
- visit houses where people refused to be interviewed to try and persuade them to cooperate;
- check on all dwellings reported by enumerators as vacant to ascertain if they were indeed vacant, and
- carry out spot checks by visiting enumerated households to verify the quality of the enumerators' work.

In addition, the supervisors had to complete several control and reporting forms to monitor movement of the questionnaires, to report to management on daily progress and problems, and to assist in quality control.

The quality control procedures in particular proved onerous. Supervisors were required to complete forms reporting on the first questionnaires submitted by each enumerator, and later to visit randomly-chosen enumerated households in each EA and complete a single-page control questionnaire against which to check the enumerators' work. The latter step was intended in particular to detect enumerators who might be cutting corners or inventing data to keep up with the workload. In the event, few supervisors managed to find time to complete these tasks, in particular the extra visits. These tasks had not been tested in the pilot census, as they had not been included in the training.

The daily progress reporting was also difficult to implement. The forms proved to be unclear and too complicated, but they had not been tested in the field, as they were developed after the pilot. Lastly, control processes had been developed by both the enumeration sub-project and the data processing sub-project, and better integration during the planning stages might have helped reduce the task burden.

The work of fieldwork co-ordinators was equally difficult to fulfil. They were responsible for receiving and checking the supervisors' progress reports, and

reporting upwards on their entire area of approximately 10 supervisor units (50 EAs). Again, the reporting forms had not been tested in the field and proved cumbersome. Moreover, much of the fieldwork co-ordinators' time was taken up in sorting out staffing, money and transport problems, rather than supervising the quality and coverage of the data collection and supporting the supervisors with their workload.

Steps taken to increase coverage

In view of various difficulties experienced in the field, it was decided to extend the enumeration period into November in some parts of the country, in what was known as the 'mop-up' operation. Reasons for this decision included:

- delays with the initial enumerator payment in some areas, which hampered the start of the actual count;
- delayed recruitment and appointment of temporary staff in some areas, due in some places to unusually adverse weather conditions prior to enumeration, and in others to problems with the CAS;
- difficulties in obtaining access to high-walled properties, areas with heavy security, and commercial farms;
- difficulty in finding people at home during the day, and insufficient arrangements for enumerating after hours.

With regard to the last two points, it was intended that enumeration staff should reflect the South African population as a whole. Where possible, enumerators should be from the community concerned, to facilitate communication and co-operation. However, insufficient enumerators were recruited from certain communities and areas. For example, few white people applied to be enumerators, either because the advertisements did not reach them, or because the money offered did not attract them. This caused problems, as enumerators working in predominantly white areas often had long distances to travel from home, which made it more difficult for them to enumerate in the evening.

In view of all the above, the extension of the enumeration period was an important strategic decision, although it was not carried out in a very systematic way. Control of the fieldwork during this extra time was less than adequate, especially as some personnel had been paid off and had already left the employ of Stats SA.

A call centre was also in operation where members of the public could call in to report that they had not been enumerated. Arrangements were then made for them

to be visited. Collection boxes were placed in post offices throughout the country for self-completed questionnaires.

All these measures helped to increase the extent of coverage.

Payment issues

Payment of an allowance to enumerators before starting the count

During the planning stages, Stats SA decided to pay an allowance to enumerators before they started the counting process, to cover costs of food and travel during enumeration.

Many problems were experienced in this regard, including incomplete lists of names being provided to the post offices. The latter, for their part, were often inadequately prepared. As a result, large amounts of cash had to be withdrawn by regional census managers to enable the initial payments to be made.

In some areas this problem delayed the start of the actual count, as noted above.

Final payment of enumerators

- The names of all enumerators on the CAS were given to a commercial bank, which provided cheques for their final payment.
- On completion of questionnaires in a given EA, these were put into the relevant bar-coded box. Both the box and the census summary book were handed to the supervisor. Questionnaires in the box were checked against the census summary book.
- The supervisor then sealed the box to prevent loss of questionnaires.
- In exchange for a completed box of questionnaires and a census summary book, each enumerator received a cheque as final payment for the work.

Handling the sealed boxes

- The sealed boxes were sent back to the regional offices, from where they were collected and delivered to the data processing centre in Pretoria.
- At the processing centre, the barcodes on each box were read and each box was allocated a specific shelf area for storage.
- All movement of boxes into and out of the storage area for processing was subsequently controlled electronically.

Chapter 4: Data processing

Introduction

The development of the data processing system began in earnest after the pilot census. This fell naturally into four phases – data capture, post-data capture, coding, and product generation.

The primary purpose of the data capture system was to transfer data from the census questionnaires to usable computer data files. It had been decided that this would be done using scanning and optical character recognition technology, and the design and development of this was outsourced to a private contractor.

The function of the post-capture system was to clean and edit the output of the data capture system and make it ready for use in a variety of census products. These subsystems were developed in-house by Stats SA.

The coding system was required to deal with open-response questions such as country of birth and religion. A computer system to assist the coding clerks in this task was developed by the private contractor.

Data capture

The backbone of the contractor's system was commercial software that had been used successfully in various census projects (such as Kenya, Turkey, Brazil and India). However, the contractor also had to develop additional software to:

- manage boxes received from the field and checked into the store;
- control and track the processing of questionnaires;
- consolidate the data received from the commercial image-processing software into database records for every questionnaire; and
- perform quality checks on the data before it was made available to Stats SA for final editing and analysis.

A business plan was compiled for processing the data from the pilot census. This defined the procedures, basic functionality and resources that would be required, as well as the budgets, time-scales and responsibilities of each party. The plan was revised and finalised after the pilot census.

The data processing of the pilot census was carried out on the Stats SA premises. For the main census, a new site in Pretoria was acquired and equipped. This had to accommodate more than 100 000 boxes of questionnaires, scanners, computer

workstations, servers and other peripheral equipment. Shelving and equipment had to be installed so that by the time boxes arrived from the field, they could be checked into the stores and processed. Two data-processing centres in Bloemfontein and Port Elizabeth were also acquired and equipped. These centres were used for data processing only and did not handle any questionnaires.

Approximately 1000 temporary personnel were employed to work in three shifts a day throughout the week, including weekends, for almost a year and a half (compared with the more than 5000 used for the 1996 Census). The contractor initially trained managers, supervisors and Stats SA trainers, who in turn trained the various operators and data-processing clerks.

The basic data capture processes comprised the following:

- Preparing and scanning 10 249 185 questionnaires received from 75 794 EAs. In total, more than 117 million images were scanned.
- Using commercial image processing and character recognition software to extract the handwritten characters from the scanned images.
- Validating and correcting extracted data by using approximately 600 operators rotated over the different shifts.
- Performing independent data quality checks by typing and comparing the data values of randomly selected field values. A target accuracy of 97% had been set, but the accuracy finally achieved was 98,9%;
- Transferring the processed values of every questionnaire to Stats SA for post-capture processing.

On 6 November 2001 the first boxes of questionnaires arrived, and data capture began in January 2002. The initial plan was to complete capture by August 2002. Unfortunately, technical difficulties intervened and production at all three centres was therefore briefly stopped mid-year and several key processes improved. Data capture was restarted on 8 July 2002 and completed by the end of February 2003.

Some of the problems and solutions included the following:

- The software developed for data processing had not been tested to the limit by the pilot, and these programmes proved unable to handle the large quantities of data from the full census. As a result, systems were developed, tested and improved under live conditions.
- Because the questionnaires were printed with a light blue background and pencils were used, the scanned images were often indistinct. This was solved by replacing the white lights in the scanners with blue lights, and all unclear images were

rescanned. In addition, a software programme was specially written for automatic image quality assurance.

- Improved monitoring procedures were developed to ensure that operators produced work of the highest possible quality.

Coding

Eight open-response (free format) questions were asked in the census: country of birth, country of citizenship, place of usual residence, place of previous residence, place of work, religion, occupation and industry. Response rates for these questions varied from relatively low (for example, country of citizenship was asked of non-SA citizens only), to very high (for example, religion was asked of every person). Each of these responses had to be linked to the standard codes used by Stats SA for countries, place names, religions, occupations and industries.

The coding task was complicated by the variation in spellings, the different languages used, and the fact that in many cases respondents used uncommon words or abbreviations. Most coding was done manually by trained coding personnel that specialised in the different open-response categories. However, the coding task for place names and religions was too big to be completed manually in the reduced time allowed and software was therefore written for automated coding of place names and religion. Approximately 60% of place names and 55% of religion responses were eventually coded automatically.

It was originally planned to complete coding over six months and publish these results after the main census results. In March 2003 a decision was taken to delay the release of the first census results and include the coded results. This decision reduced the time allowed for coding from six months to ten weeks. Coding finished mid-May 2003.

Certain difficulties were experienced, but most were solved or satisfactory compromises achieved:

- Due to the greatly reduced time-frame, additional coding staff had to be appointed. Coding is a highly specialised field requiring experts in each category. Unfortunately, not all coding staff could be trained in time to achieve the same results as professional coders.
- Insufficient space on the questionnaires prevented accurate coding for some of the responses. In addition, unclear responses were sometimes received and these were difficult to code.

- The dictionaries used were incomplete and sometimes place names were used that were not on the list. A lack of time made it virtually impossible to resolve all such cases.
- The initial plan of coding to a sub-place level presented considerable difficulties and it was agreed to code place names to a main-place level only.

Post-capture processing

After the data was delivered to Stats SA, a series of post-capture processes took place, including:

- removing erroneous person and death records created when the scanning system misinterpreted blank fields;
- correcting errors caused when the enumerator skipped up or down a line between pages while recording fertility responses, and
- resolving problems with large households, where the enumeration required more than a single questionnaire.

Following these post-capture phases, the raw data was analysed and edited. A set of automated editing specifications was developed. A team of demographers, subject-matter specialists and computer programmers elaborated and implemented the editing system. Derived variables were developed at this stage of processing.

For the first time, Stats SA used imputation to correct missing or inconsistent values. In such situations, a reasonable valid value was substituted for the problematic one. A combination of logical imputation methods (based on the characteristics of other members of the household) and dynamic imputation methods (based on characteristics shared with people in other nearby households) was used.

The result of post-capture processing was clean final data, ready for tabulation and dissemination. Editing was completed less than one month after the end of coding, and final data sets were delivered to Stats SA management in early June 2003.

Summary

The Census 2001 questionnaire was detailed and complex, and this made the process of editing the census data extremely challenging. As an example, the questionnaire contained variables for both age and date of birth, and the editing system had to ensure that these were consistent. If they were not, a series of tests was

performed to decide which response was more likely to be correct, and the other was imputed to make it consistent. Some questions, such as 'spouse person number' required elaborate specifications to fully edit. Such specifications took substantial time to develop and test.

The editing system had to be adjusted in April 2003 in order to accommodate the revised release schedule and include open-response coded variables. This required adjustments to many of the editing specifications and associated computer systems.

PLEASE SCROLL DOWN

Chapter 5: The post-enumeration survey

A post-enumeration survey (PES) is undertaken to determine the degree of under- or overcount in a population census, and to evaluate the quality of data collected during the census. In Census 2001, this exercise was undertaken shortly after the enumeration phase of the census, during November and early December 2001. The goal is to conduct an exhaustive, independent enumeration in the selected EAs to compare against census data.

There are two basic types of error that may be identified by a post-enumeration survey:

- errors of coverage, which include both erroneous omissions and erroneous inclusions. An example of an erroneous omission is a household that was not counted in the census because the enumerator did not visit it. An example of erroneous inclusion is a household that moved into the EA after census night and was still counted when the enumerator visited; and
- errors of content, which concern the reported characteristics of the people or households enumerated.

The errors may be expanded as follows:

- failure to account for all inhabited areas in the EA frame;
- EA boundary problems;
- incomplete listing of dwellings within the EA (failure to identify all dwellings);
- failure to visit all listed dwellings;
- failure to identify all households, where multiple households are found in dwellings;
- failure to complete questionnaires for all households (refusals, non-return of questionnaires left for self-enumeration, etc.);
- failure to include all individuals within households;
- failure to observe the inclusion rule based on a person's presence on census night, in other words failure to apply the *de facto* census accurately;
- lost questionnaires or questionnaires that could not be processed.

Methodology

In line with the procedure followed during the census, the 2001 PES involved identifying the boundaries of each EA in the sample, listing all the dwellings within the EA, and administering a questionnaire to each household. The PES questionnaire gathered some details of the members of the household as a check upon the accuracy of the census information.

Conducting the PES involved the following stages:

- preparation, consisting of drawing a sample, questionnaire development, and staffing and training;
- fieldwork, consisting of re-mapping on the ground, re-listing of dwellings, questionnaire administration, and completing a new census summary book;
- matching the households in each EA visited during the PES, and the individuals within them, to households and individuals in the same EAs in the census;
- carrying out reconciliation visits in the field in cases where matching was uncertain;
- data capture and clean-up; and
- calculation of coverage and content analysis measures, including the extent of the undercount.

Preparation

Sampling

A sample of 600 EAs was drawn for the PES in the following manner:

- The country was stratified by province, as well as by urban/non-urban classification (four types).
- Independent systematic samples of EAs were then drawn for each stratum within each province with equal probabilities.
- The sampling frame used for the PES was constructed from the preliminary database of EAs established during the demarcation and listing phase of the census.
- Empty (consisting entirely of vacant sites), industrial and institutional EAs were excluded from the sample.

The questionnaire

The questionnaire used for the PES was a substantially shorter version of the census questionnaire. It sought to establish a list of all persons present in the household on the census reference night as well as all those present on the PES reference night, with the whereabouts of each person on both nights. It included questions on age, sex, marital status, population group, home language and level of education for each person listed.

PES enumeration

Altogether, approximately 1 500 temporary staff, including fieldworkers or enumerators (two in each EA), fieldwork supervisors and provincial managers, were employed to work on the PES.

Most of the staff was selected from household survey fieldworkers who had been recommended by their managers as being highly competent in carrying out their interviews and performing associated tasks.

Stats SA Head Office staff trained provincial managers for the PES. The provincial managers in turn trained the fieldwork supervisors and fieldworkers at provincial level. Training manuals were developed for all three levels of staff. The aim was to achieve uniform and reliable enumeration in the PES.

The fieldworkers were required to work in pairs on the following tasks:

- Check the boundaries of the EA to which they were assigned. New summary books (PES summary books), containing EA maps and orientation maps, were supplied.
- List afresh all the dwellings in the EA, without having access to the previous listings.
- Administer questionnaires in each household and hostel in the EA.
- Complete the PES summary book in the same way as in the census, (1) by filling in the number of males and females, and the total number of people, initially for each questionnaire completed and ultimately for the whole EA; and (2) by indicating all vacant premises or stands, premises which contained non-residential buildings (such as churches, schools and business premises), households who refused to be interviewed and households who could not be contacted.

The matching process

The PES data processing plan was originally based on automated computer-assisted matching. Due to initial problems with the scanning process, and the delays in implementing the new scanning procedures caused by the re-design of the system, a manual matching system was introduced. The objective of the matching exercise was to compare the PES questionnaires with their equivalents in the census.

The following is a brief description of the way in which matching was done. For a more detailed description, the reader should consult the PES methodology report, available on the Stats SA website.

Before matching began, each person was classified as one of the following:

- a non-mover (someone who was present in the same household both on census night and on PES night);
- an in-mover (someone who was not present in the household on census night, but who was present on PES night);
- a person who was born after census night; or
- an out-mover (someone who was present in the household on census night but was absent on PES night).

Initial matching

The initial matching produced one of seven possible results for each person:

- matched – the same person was found in both the census and the PES;
- possible match – there were unexplained differences between the two databases;
- in the PES but not in the census because:
 - a) there was a definite non-match;
 - b) there was insufficient or unclear information;
 - c) the person was an in-mover; or
 - d) the person was born after census night; or
- in the census but omitted from the PES.

Reconciliation visits

Cases of 'possible match', 'in the PES but not in the census because of insufficient or unclear information', and 'in the census but not in PES' were identified for reconciliation visits. In addition, reconciliation visits were carried out in all EAs with boundary problems or with overall poor quality results. Special forms were developed for these visits.

Final matching

After the reconciliation visits, the possible outcomes for final matching were as follows:

- matched;
- in the PES and not in the census, because:
 - a) missed in the census;
 - b) included erroneously in the PES;

- c) insufficient information in the PES;
- d) in-mover; or
- e) born after the census; or
- in the census and not in the PES, because:
 - a) correctly enumerated in the census but missed in the PES;
 - b) included erroneously in the census;
 - c) insufficient information in the census;
 - d) out-mover; or
 - e) died since the census.

The adjustment process

The following table indicates the undercount rates as established by the PES.

Table 5.1: Percentage undercount for persons and households in each province

Province	Persons	Households
Eastern Cape	14,74	15,55
Free State	17,63	20,60
Gauteng	18,74	23,02
KwaZulu-Natal	22,51	26,21
Limpopo	14,36	17,04
Mpumalanga	16,08	17,24
Northern Cape	14,07	17,81
North West	16,02	20,29
Western Cape	16,27	16,93
South Africa	17,64	20,52

The actual adjustment procedure consisted of creating homogeneous adjustment classes with similar coverage rates and calculating a common undercount rate, adjustment factor and adjustment figure for each class separately. The adjusted figure for the total population was obtained by summing across the adjustment classes. In addition, only the population of households and hostels received adjustment factors. The totals for the balance of the population, namely people living in all collective living quarters other than hostels, and the homeless on the street, were not adjusted.

PLEASE SCROLL DOWN

Chapter 6: Key results

The processing and weighting of the data collected during Census 2001 is now complete and Stats SA can announce the results.

The population of South Africa

On the night of 9 to 10 October 2001 there were 44,8 million people in South Africa. This total has been adjusted for undercount, using the findings of a post-enumeration survey (PES).

The numbers and percentages presented are based on empirical evidence. Both the census count and PES numbers were arrived at by visiting private dwellings and communal living quarters throughout the country, and obtaining information on the people living in them.

Population by province and sex

Of the total of 44,8 million people, 21,4 million (47,8%) were male and 23,4 million (52,2%) were female. Table 6.1 gives the number and percentage of the male and female population in each province. The percentages add up to 100 across each row. For example, if we look across the row for the Eastern Cape, column (b) shows that 46,2% of people in the province were male, while column (d) shows that 53,8% were female, adding up to a total of 100% in column (f).

Table 6.1: The population of South Africa by province and sex

Province	Male		Female		Total	
	N (a)	% (b)	N (c)	% (d)	N (e)	% (f)
Eastern Cape	2 975 512	46,2	3 461 251	53,8	6 436 763	100,0
Free State	1 297 605	47,9	1 409 170	52,1	2 706 775	100,0
Gauteng	4 444 679	50,3	4 392 499	49,7	8 837 178	100,0
KwaZulu-Natal	4 409 091	46,8	5 016 925	53,2	9 426 017	100,0
Limpopo	2 394 785	45,4	2 878 857	54,6	5 273 642	100,0
Mpumalanga	1 497 333	47,9	1 625 658	52,1	3 122 990	100,0
Northern Cape	401 168	48,8	421 559	51,2	822 727	100,0
North West	1 821 547	49,6	1 847 803	50,4	3 669 349	100,0
Western Cape	2 192 321	48,5	2 332 014	51,5	4 524 335	100,0
South Africa*	21 434 040	47,8	23 385 737	52,2	44 819 778	100,0

* All PES-adjusted numbers are rounded to whole numbers. The totals displayed may therefore differ slightly from the sum of the separate numbers.

The table shows that in only one province, namely Gauteng, were there proportionately more men (50,3%) than women (49,7%). It also shows that Limpopo and the Eastern Cape had the highest proportion of women (54,6% and 53,8% respectively) and the lowest proportion of men (45,4% and 46,2%) compared with the other provinces. The internal migration of young men from the more rural to the more urban provinces may in part explain this finding.

Population by province and population group

As many as 35,4 million (79,0%) of the South African population were African, and this population group was in the majority in seven of the nine provinces.

Table 6.2 gives the number and percentage of people living in each province by population group. The percentages add up to 100 across the rows.

- It shows that 97,2% of people living in Limpopo and 91,5% of people living in North West were African.
- On the other hand, the coloured population group was the largest group in the Western Cape (53,9%) and the Northern Cape (51,6%).

Table 6.2: The population of South Africa by population group within provinces

Province (a)	Black African		Coloured		Indian or Asian		White		Total	
	N (b)	% (c)	N (d)	% (e)	N (f)	% (g)	N (h)	% (i)	N (j)	% (k)
Eastern Cape	5 635 079	87,5	478 807	7,4	18 372	0,3	304 506	4,7	6 436 763	100,0
Free State	2 381 073	88,0	83 193	3,1	3 719	0,1	238 791	8,8	2 706 775	100,0
Gauteng	6 522 792	73,8	337 974	3,8	218 015	2,5	1 758 398	19,9	8 837 178	100,0
KwaZulu-Natal	8 002 407	84,9	141 887	1,5	798 275	8,5	483 448	5,1	9 426 017	100,0
Limpopo	5 128 616	97,2	10 163	0,2	8 587	0,2	126 276	2,4	5 273 642	100,0
Mpumalanga	2 886 345	92,4	22 158	0,7	11 244	0,4	203 244	6,5	3 122 990	100,0
Northern Cape	293 976	35,7	424 389	51,6	2 320	0,3	102 042	12,4	822 727	100,0
North West	3 358 450	91,5	56 959	1,6	9 906	0,3	244 035	6,7	3 669 349	100,0
Western Cape	1 207 429	26,7	2 438 976	53,9	45 030	1,0	832 901	18,4	4 524 335	100,0
Total*	35 416 166	79,0	3 994 505	8,9	1 115 467	2,5	4 293 640	9,6	44 819 778	100,0

* All PES-adjusted numbers are rounded to whole numbers. The totals displayed may therefore differ slightly from the sum of the separate numbers.

Table 6.3 below shows the distribution of each major population group by province. The numbers are the same as in Table 6.2, but the percentages are different, since they add up to 100 down each column, e.g. the third column from the left indicates that 15,9% of Africans lived in the Eastern Cape, while 6,7% lived in the Free State.

The table shows that Africans and whites were more widely dispersed across the country than coloureds and Indians/Asians.

- Relatively large proportions of the total African population lived in KwaZulu-Natal (22,6%), Gauteng (18,4%), the Eastern Cape (15,9%) and Limpopo (14,5%) while small proportions lived in both the Western Cape (3,4%) and the Northern Cape (0,8%). This distribution may reflect previous labour policies, where work preference was given to coloured rather than African people in these regions.
- The coloured population was found mostly in three provinces: the Western Cape (61,1%), the Eastern Cape (12,0%) and the Northern Cape (10,6%). Only small proportions of coloured people lived in Limpopo (0,3%) and Mpumalanga (0,6%).
- More than seven-tenths of the Indian/Asian population were found in KwaZulu-Natal (71,6%), with the second-largest proportion (19,5%) being found in Gauteng. Only small proportions lived in the Northern Cape (0,2%), Free State (0,3%) and Limpopo (0,8%).
- The white population tends to live in those provinces that are largely urbanised. The largest proportion of white people was found in Gauteng (41,0%), followed by the Western Cape (19,4%) and KwaZulu-Natal (11,3%).

Table 6.3: The population of South Africa by province within each population group

Province (a)	Black African		Coloured		Indian or Asian		White		Total	
	N (b)	% (c)	N (d)	% (e)	N (f)	% (g)	N (h)	% (i)	N (j)	% (k)
Eastern Cape	5 635 079	15,9	478 807	12,0	18 372	1,6	304 506	7,1	6 436 763	14,4
Free State	2 381 073	6,7	83 193	2,1	3 719	0,3	238 791	5,6	2 706 775	6,0
Gauteng	6 522 792	18,4	337 974	8,5	218 015	19,5	1 758 398	41,0	8 837 178	19,7
KwaZulu-Natal	8 002 407	22,6	141 887	3,6	798 275	71,6	483 448	11,3	9 426 017	21,0
Limpopo	5 128 626	14,5	10 163	0,3	8 587	0,8	126 276	2,9	5 273 642	11,8
Mpumalanga	2 886 345	8,1	22 158	0,6	11 244	1,0	203 244	4,7	3 122 990	7,0
Northern Cape	293 976	0,8	424 389	10,6	2 320	0,2	102 042	2,4	822 727	1,8
North West	3 358 450	9,5	56 959	1,4	9 906	0,9	244 035	5,7	3 669 349	8,2
Western Cape	1 207 429	3,4	2 438 976	61,1	45 030	4,0	832 901	19,4	4 524 335	10,1
Total*	35 416 166	100,0	3 994 505	100,0	1 115 467	100,0	4 293 640	100,0	44 819 778	100,0

* All PES-adjusted numbers are rounded to whole numbers. The totals displayed may therefore differ slightly from the sum of the separate numbers. Similarly, the percentages are rounded to the first decimal place, and therefore may not always add up to exactly 100.

The age distribution

The age distribution by five-year intervals is given in Table 6.4 below. It can be seen in column (c) that 32,1% of people in South Africa were under the age of 15, while 4,9% were aged 65 or older. This distribution reflects that, in general, the country has a relatively young population. However, these overall patterns mask differences in age distribution within the country not only by sex, but also among the different population groups.

Table 6.4: Age distribution of South Africans in five-year intervals

Age group (a)	N (b)	% (c)	% (d)	Sex ratio* (e)
85+	157 333	0,4		41,2
80-84	270 945	0,6		50,4
75-79	367 537	0,8	4,9	59,0
70-74	631 469	1,4		58,3
65-69	787 927	1,8		63,1
60-64	1 065 294	2,4		71,6
55-59	1 205 266	2,7		84,6
50-54	1 638 020	3,7		88,6
45-49	2 087 380	4,7	42,3	86,4
40-44	2 619 465	5,8		89,0
35-39	3 071 770	6,9		88,4
30-34	3 340 901	7,5		91,3
25-29	3 934 939	8,8		93,3
20-24	4 294 523	9,6	20,7	95,6
15-19	4 981 721	11,1		97,0
10-14	5 061 917	11,3		99,1
5-9	4 853 555	10,8	32,1	99,9
0-4	4 449 816	9,9		99,9
Total	44 819 778	100,0	100,0	91,7

* No. of men for every 100 women.

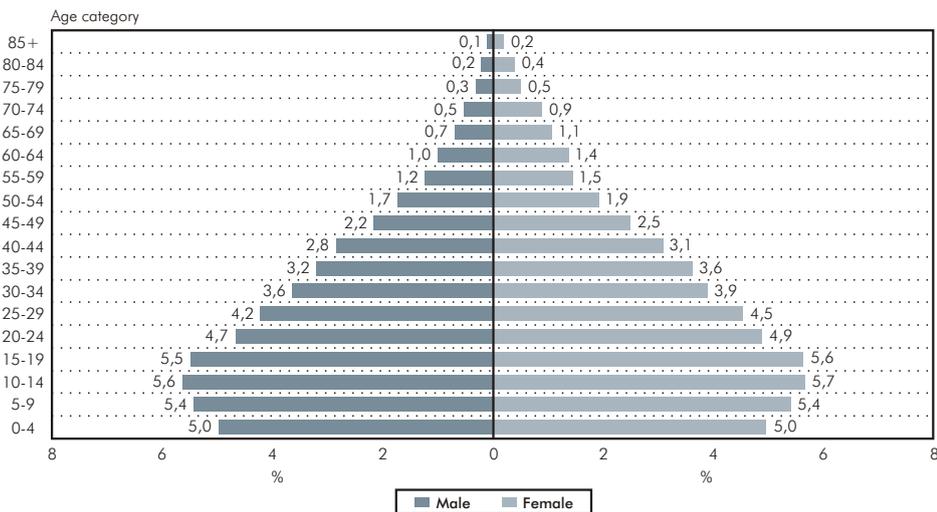
All PES-adjusted numbers are rounded to whole numbers. The totals displayed may therefore differ slightly from the sum of the separate numbers. Similarly the percentages are rounded to the first decimal place, and therefore may not always add up to the totals displayed.

The sex ratios (the number of males for every 100 females, as shown in column (e) of Table 6.4) were lower than expected in certain age categories, for example, the 0–4 year age group and young adults (aged 20–34). This may be due to under-reporting of babies, and the difficulty in reaching young single males during a census.

Examining sex differences in age distribution, it can be seen in Figure 6.1 that of the total population, 16,0% were males under the age of 15, and 16,1% were females under the age of 15 – slightly more females than males. At the other end of the

scale, 1,9% were males over the age of 64, whilst 3,2% were females over the age of 64, indicative of a longer life expectancy among females than males. (Note that in all the age pyramids, the whole graph, i.e. males and females, adds up to 100%.)

Figure 6.1: Distribution of the total population by age group and sex



Figures 6.2 to 6.5 give the age distribution in five-year intervals for each of the major population groups in the country. Differences in the age distribution of each population group show various stages in the transition from a younger to an older population.

- Among Africans, the overall pattern shows that there were relatively large proportions of males and females below the age of 15 (17,0% and 17,1% respectively of the total population), and relatively small proportions aged 65 or more (1,4% males and 2,8% females).
- Among the coloured population group, smaller proportions than of Africans were under the age of 15 (15,5% males and 15,3% females), and smaller proportions of both males (1,5%) and females (2,5%) were aged 65 or older.
- Among the Indian/Asian group, an even smaller proportion (23,5%) was below the age of 15.
- The white population group had the smallest proportion of all aged less than 15 (19,0%). However, a far larger proportion than of all the other groups was aged 65 or older (4,7% males and 6,5% females).

Figure 6.2: Distribution of the black African population by age group and sex

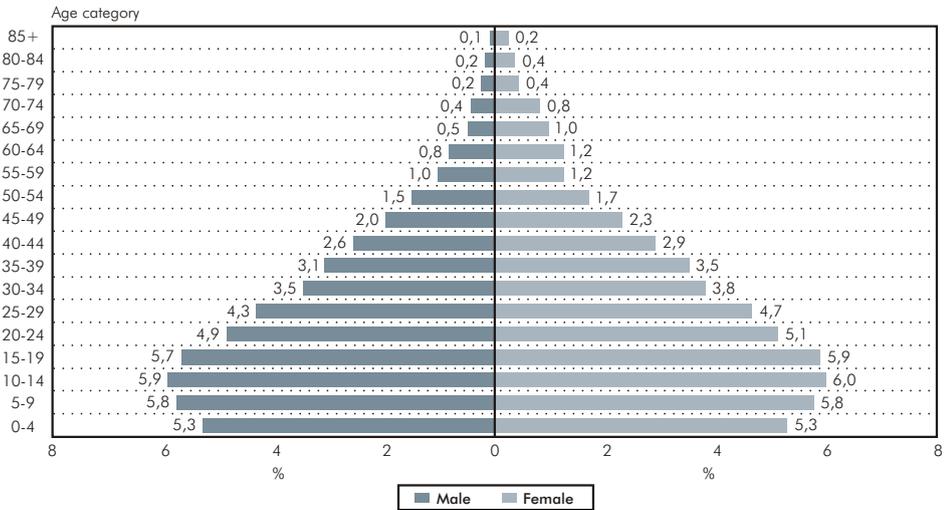


Figure 6.3: Distribution of the coloured population by age group and sex

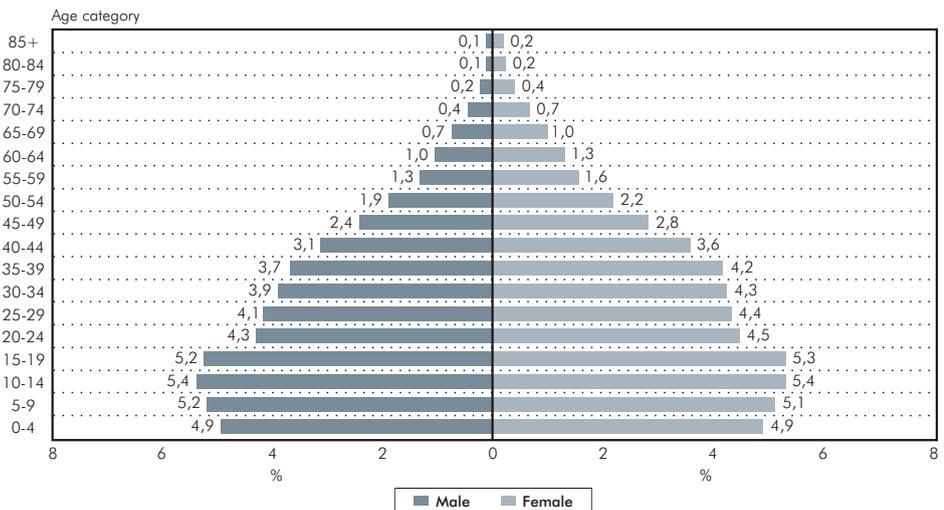


Figure 6.4: Distribution of the Indian and Asian population by age group and sex

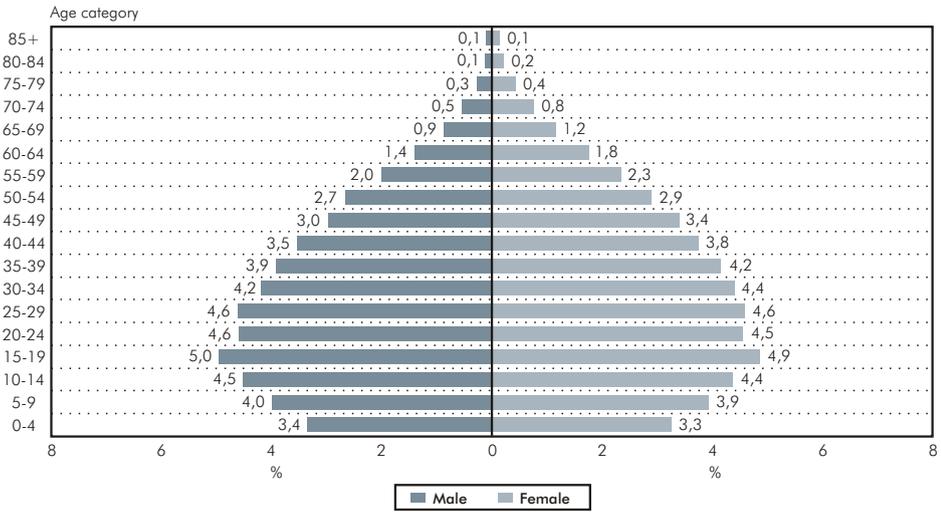
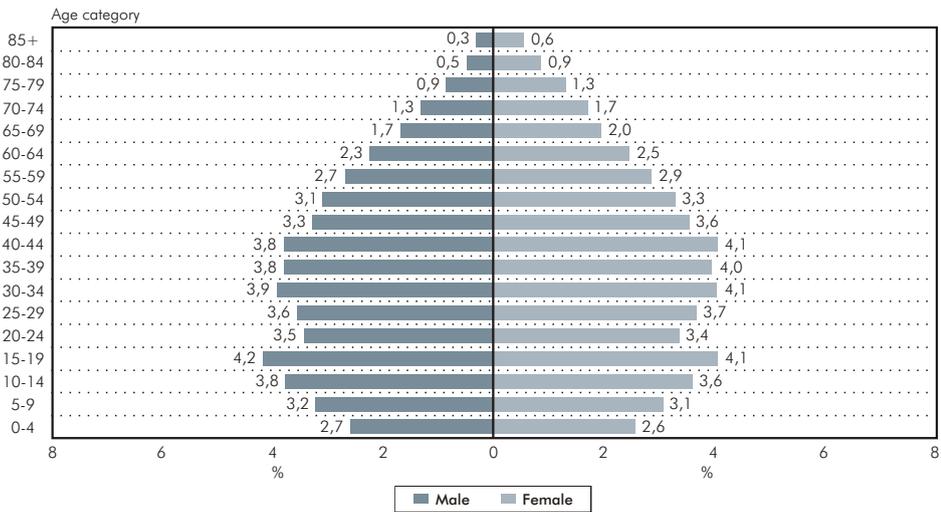


Figure 6.5: Distribution of the white population by age group and sex



The households of South Africa

There were 11,2 million households in South Africa in October 2001. This compares with 9,0 million in 1996. (This figure does not include households or individuals living in institutions, hostels, or other collective living quarters.) Table 6.5 indicates the type of dwelling in which these households were living by population group of the household head. The percentages add up to 100 down each column. For example, among African-headed households (column (b)), 55,5% were living in formal dwellings such as a house or a flat, 20,4% were living in informal dwellings such as shacks, and 18,7% in traditional dwellings.

- Relatively few African-headed households were living in formal dwellings, compared to households headed by coloured, Indian/Asian or white people.
- A relatively large proportion of African-headed households, compared with the other population groups, were living in traditional dwellings or shacks.

Table 6.5: Households in South Africa by type of dwelling and population group of the household head

Type of dwelling (a)	Black African		Coloured		Indian or Asian		White		Total	
	N (b)	% (c)	N (d)	% (e)	N (f)	% (g)	N (h)	% (i)	N (j)	% (k)
Formal	4 783 620	55,5	761 115	85,7	262 386	92,7	1 340 319	95,1	7 147 438	63,8
Informal	1 760 410	20,4	65 833	7,4	3 041	1,1	6 947	0,5	1 836 232	16,4
Traditional	1 610 402	18,7	24 967	2,8	3 993	1,4	15 424	1,1	1 654 787	14,8
Backyard formal	444 994	5,2	33 098	3,7	12 888	4,6	42 002	3,0	532 983	4,8
Other	25 624	0,3	3 023	0,3	621	0,2	4 997	0,4	34 266	0,3
Total*	8 625 050	100,0	888 036	100,0	282 929	100,0	1 409 689	100,0	11 205 705	100,0

Excluding households in collective living quarters.

* All PES-adjusted numbers are rounded to whole numbers. The totals displayed may therefore differ slightly from the sum of the separate numbers. Similarly the percentages are rounded to the first decimal place, and therefore may not always add up to exactly 100.

Summary of adjustments

Table 6.6 compares the numbers of persons and households actually counted with the final estimates, by province and in the country as a whole. For example, in the Eastern Cape, 5,5 million people were actually counted in the census. On the basis of the findings of the PES, when the raw count was adjusted to include those who were missed, the final estimate is 6,4 million. The table reads similarly for the households.

More details of unadjusted and adjusted figures, and 95% confidence limits, may be found in Appendix B.

Table 6.6: Census count and adjusted count of the population and households, by province

Province	Persons		Households	
	Unadjusted count	Adjusted count	Unadjusted count	Adjusted count
Eastern Cape	5 537 841	6 436 763	1 288 456	1 512 664
Free State	2 255 442	2 706 775	587 518	733 302
Gauteng	7 270 597	8 837 178	2 079 100	2 651 244
KwaZulu-Natal	7 392 274	9 426 017	1 572 591	2 086 250
Limpopo	4 543 051	5 273 642	1 000 619	1 179 965
Mpumalanga	2 641 152	3 122 990	617 505	733 131
Northern Cape	714 708	822 727	172 870	206 842
North West	3 108 050	3 669 349	753 410	929 004
Western Cape	3 839 068	4 524 335	985 411	1 173 304
South Africa*	37 302 183	44 819 778	9 057 480	11 205 705

Excluding households in collective living quarters.

* All PES-adjusted numbers are rounded to whole numbers. The totals displayed may therefore differ slightly from the sum of the separate numbers.

KwaZulu-Natal has the largest population (9,4 million), followed by Gauteng (8,8 million). The Northern Cape has the smallest population (0,8 million) and the Free State the second smallest (2,7 million).

Gauteng has the largest number of households in the country (2,7 million), even though it does not have the largest population. The Northern Cape has the smallest number of households (0,2 million).

Appendix A: Members of the Statistics Council

Members of the Census sub-committee are indicated with an asterisk.

Mrs LNN Bhengu-Baloyi
Mr LC Fouché*
Prof RE Dorrington*
Prof J Galpin* (Chair of Census sub-committee)
Mr MJ Jack
Dr MA Lesaoana
Dr C Loewald
Mr JM Mamabolo
Ms M Marais-Martin
Prof JD May*
Mr M McDonald
Dr CE Meth
Mr N Mokhesi*
Mr CM Morolo
Prof TZ Mthembu
Mr R Naidoo
Mr FY Patel
Prof TB Pretorius
Mr JW Prinsloo
Prof CEW Simkins*
Dr HA Southall* (Chair of Council)
Mr GE Thabane

Appendix B: PES adjustment

Adjusted population figures by province, population group, sex and age, with 95% confidence limits

Category	95% Confidence Interval Limits		
	Estimate	Lower	Upper
All persons	44 819 778	44 427 683	45 211 872
Province			
Eastern Cape	6 436 763	6 286 402	6 587 125
Free State	2 706 775	2 665 303	2 748 247
Gauteng	8 837 178	8 520 018	9 154 338
KwaZulu-Natal	9 426 017	9 030 906	9 821 128
Limpopo	5 273 642	5 244 376	5 302 907
Mpumalanga	3 122 990	3 081 917	3 164 064
Northern Cape	822 727	812 071	833 384
North West	3 669 349	3 608 191	3 730 507
Western Cape	4 524 335	4 439 010	4 609 661
Population group			
Black African	35 416 166	34 923 119	35 909 213
Coloured	3 994 505	3 917 140	4 071 871
Indian or Asian	1 115 467	1 084 589	1 146 345
White	4 293 640	4 205 194	4 382 086
Sex			
Male	21 434 040	21 182 666	21 685 415
Female	23 385 737	23 108 636	23 662 839
Age group			
0-4	4 449 816	4 390 734	4 508 897
5-14	9 915 472	9 768 819	10 062 125
15-19	4 981 721	4 920 430	5 043 011
20-29	8 229 462	8 133 430	8 325 494
30-44	9 032 136	8 925 942	9 138 330
45-64	5 995 960	5 930 696	6 061 224
65+	2 215 211	2 191 652	2 238 771

Appendix C: Census products

The development, dissemination and marketing of products are ongoing processes. They include deciding on the range of products and services to be offered, the software to use, the pricing structures, and what will be made available free of charge.

The overall objective is to produce and disseminate user-friendly products and services targeted to different audiences. A variety of census products will be distributed in both print and electronic format. Libraries, schools, community organisations, business organisations, trade unions and government departments countrywide will receive print copies of certain products. These will also be made available to the general public from Stats SA's head office and its provincial offices. Electronically, a series of databases will be made available on the Internet.

Print products

The following print products are now available:

- *Key census results*, a pamphlet aimed at the general public, which outlines briefly how the count was done and contains a few highlights of the results (Ref: 03-02-01).
- *Census in brief*, an A6 booklet consisting of over 80 tables and graphs at national and provincial level, for an extensive range of individual and household variables (Ref: 03-02-03).
- This report *How the count was done* (Ref: 03-02-02).
- Thematic and other posters.

Other print products will be made available in due course, including:

- *Primary tables*, giving more detailed information on the results in tabular form, for the country as a whole (Ref: 03-02-04) and for each province (Ref: 03-02-05 to 03-02-13).
- *Other written reports*, including a summary report (Ref: 03-02-16), in-depth reports on particular themes, additional methodology reports including the post-enumeration methodology (Ref: 03-02-17), and the census review (Ref: 08-02-18).
- *Key municipal data*, which contains breakdowns at municipal level for a range of individual and household variables (Ref: 30-02-21).

All the print products will be made available on the Internet when they are ready.

Electronic products

The following electronic products will be available on the Internet:

- *Interactive Internet products*: a series of interactive products, for users to compile tables according to their own specifications.
- *Community profiles*: for users who wish to arrange and combine information into their own unique tables, at different levels of geography (Ref: 03-02-22).
- *Investigation into the definitions of 'urban'*. This discussion document is Stats SA's first attempt to get to grips with the problems involved in categorising areas as urban or rural. The paper identifies differences in this regard between Census '96 and Census 2001 and examines the census results in the light of these differences. It then experiments with a definition of 'urban' based on population density. Two versions of the paper are available, a summary and a detailed report. Both contain tables and colour maps, and can be downloaded from the web in PDF format.
- *Age tables* by single-year breakdowns for the country as a whole (Ref: 03-02-30) and for each province (03-02-31 to 03-02-39).
- *Census concepts and definitions*: an alphabetical listing of concepts and definitions used during the census, with some methodological notes (Ref: 03-02-26).
- *Other general and geographical metadata files*. The general metadata files include, among other things, the exact wording of each question, the guidelines that were given to the enumerators on how to interpret the replies, and the final code lists for all census data (Ref: 03-02-24). The geographical metadata file explains the geography of the census and the coding of all the geographic areas in the country (Ref: 03-02-25).
- *Demographic atlas*. This product will display the demographic characteristics of various towns, cities and municipalities (Ref: 03-02-28).

In addition, the following electronic products will be available on request:

- CD Rom disks containing the *community profiles* described above (Ref: 03-02-22).
- *My constituency*: a CD with census findings for all electoral wards, designed to give parliamentarians more information about the wards they represent (Ref: 03-02-29).
- *Special requests*. A set of tables can be produced by Stats SA, either at head office or in each province, providing specific information, at any level of geography: EA, sub place, main place, municipal, magisterial district or provincial level.

- *Sample database in SuperCross and ASCII*. This sample of census records is designed for researchers wishing to do their own analyses. (Ref: 03-02-23).
- *CD with spatial (GIS) data*. This product is designed for users with their own GIS software. These digitised enumeration areas and boundaries can serve as a backdrop for any GIS system. The CD contains information about all geographical areas in the country, from provincial to the smallest area (Ref: 03-02-27).

Publications can be ordered from: Printing and Distribution, Statistics South Africa

Tel: (012) 310 8251

Fax: (012) 322 3374

E-mail: distribution@statssa.gov.za