South Africa's 2022 census missed 31% of people - big data could help in future

Published: October 13, 2023 2.17pm SAST

Author

David Everatt
Professor of Urban Governance, University of the Witwatersrand

Enumerators using electronic tablets in South Africa's census on 2 February 2022. Phil Magakoe/AFP via Getty Images

No census is ever exact: as academics Tom Moultrie and Rob Dorrington at the University of Cape Town have noted previously:

a census is not, in reality, a full and accurate count of the number of people in a country; rather, it is itself an estimate of the size of the population at a moment in time.

South Africa has announced the results of its fourth census as a democracy – Census 2022. I have been involved in the process for the last four years as chair of South Africa's National Statistics Council. As outgoing chair, my last task was to take part in the release of Census 2022.

The census found that the national population has grown to 62 million, up 10.3 million from the last census in 2011. Gauteng is now clearly the most populous province in the country, with 15.1 million people, overtaking KwaZulu-Natal (12.4 million). The Western Cape jumped from fifth to being the third largest province, with 7.4 million people. These figures are important because they inform resource allocation by government.

Get your news from people who know what they're talking about.

Get newsletter

What is perhaps most striking about Census 2022 is the very high undercount – 31% of people and 30% of households were missed (or chose not to self-enumerate, either online or via zero-rated telephone methods). This is the highest undercount of any post-apartheid census; sadly, it may set a new international record.

A census is immediately followed by a Post Enumeration Survey, which identifies where the census missed people. This allows Statistics SA to develop adjustment factors, or weights, so that the final data represents an adjusted final tally. The Post Enumeration Survey is used to manage the undercount. Census undercounts are the norm, not the exception. But it is safe to assume that with weighting on this scale – adjusting for an undercount of 31.06% – analysts may identify some confounding results.

At aggregate level, Census 2022 is robust. At sub-national – and especially sub-provincial – levels, however, it may be less so. Only time and data analysis will tell.

The census confirmed the global trend of declining survey response rates. People are less and less inclined to be involved in the process. This raises the question: does a fieldwork-based census have a future? Given the challenges that faced Census 2022, I believe the census may need to be re-imagined as a very different exercise. This requires Statistics South Africa, which conducts the census, to fully engage with big data to bring the process into the 21st century.

The process

South Africa's National Statistics Council, an independent body of experts that advises the statistician-general and the minister in the presidency regarding statistics, had secured a number of local and international experts – as had Stats SA – to stress test the census and the Post Enumeration Survey. Council never has prior sight of the data: its job is to focus on methods and process.

The experts do engage with the data and flagged only a few variables (mortality data, and some service and asset questions which had too many non-responses to be reliable) as requiring a cautionary note. Council engaged vigorously with the experts and Stats SA, and with no red flag raised by any, we declared the census "fit for purpose".

It is notable that Stats SA routinely conducts a post enumeration survey. Many countries do not, even when there is systematic undercounting of particular groups (often young men, children and minorities). Moreover, Stats SA will make available both the weighted and the raw data for analysts to examine in detail. This transparency should be welcomed, given that (as previously noted by the United Nations Statistics Division) issues of undercounting affect all countries, and estimating the undercount and whether to adjust the data is a political issue "throughout the world". The undercount was high, but not as a result of any lack of effort or commitment from Stats SA.

Why the undercount

The undercount is the result of many factors.

First, the context matters. This time round it was as bad as it could be, with the COVID-19 pandemic affecting training and supply chains for equipment. The pandemic also generated anxiety in a populace that had been avoiding contact with strangers as part of social distancing. Census planning usually starts three or four years prior to fieldwork. Training about 100,000 enumerators is a major effort in its own right, combined with the shift to digital platforms for the first time. All were affected by the pandemic.

The fieldwork took place after the devastating July 2021 insurrection, and after the hard-fought local elections. The process also coincided with xenophobic violence meted out by the anti-migrant pressure group-turned-political party Operation Dudula in Johannesburg. Taken together, the effect was a deep-seated reluctance to open doors to strangers, particularly those asking lots of questions.

A second factor that affected the gathering of data was the fact that there is very low trust in the government. Although the census is conducted by Stats SA, which is an independent entity, it is seen as "government". This label didn't make it easy to persuade people to allow an enumerator into their dwellings and answer questions.

People in the Western Cape, the only province not run by the African National Congress, were particularly resistant to being enumerated or self-enumerating. This was true even after the provincial premier and Cape Town mayor made public calls for people to comply. The undercount in the Western Cape stands at 35.58% of people and 36.3% of households. In the Free State, by comparison, the undercount is 20.95% of people and 17.93% of households.

A third factor was that response rates have been getting consistently lower over at least the last decade. This has been true for Stats SA and other entities undertaking primary research. The decision to go digital was an attempt to open different avenues for people to

complete the questionnaire online, or by phone, to improve response rates.

People appear to be sick and tired of being polled by everyone, from their local supermarket to endless tele-marketers and others. They also appear much more wary of sharing their data. What, then, is the future for the census?

Enter big data

Countries around the world are facing the same challenge of low response rates.

The advent of big data opens intriguing possibilities.

A first step would be to harvest data from the records kept by government departments (assuming they are run well). In addition, data could be unlocked if a working relationship was developed with private sector entities, such as suppliers and banks.

Becoming far more tech-savvy, and encouraging people to engage with Stats SA digitally, could be combined with other options to compile a national population dataset. It would also represent a significant cost-saving. This approach – harvesting data rather than gathering it directly – is being considered by many countries, but has not yet been attempted, and Stats SA needs to carefully consider this option.

Stats SA needs to fully engage with the world of big data, and the key players in that data ecosystem. It has convening authority, and should be engaging all key players, whether they are academic, private sector or others.

At the very least, an alternative way of conducting the next census in 2032 must be rigorously examined and tested.

Big data is not the answer to all the challenges that faced Census 2022, but it may be a key enabler for gathering reliable national data in the future.