stats sa

Department:
Statistics South Africa
REPUBLIC OF SOUTH AFRICA

November 2025

Ward Level Small Area Population Estimates, 2022

Technical Note

Contents

1. Introduction

This technical note outlines processes and methods used to derive Small Area Estimates (SAEs) of the population structure at ward level, utilising intercensal Census 2011 and Census 2022 data. Small Area Estimates are estimates for area-level domains produced from mutually consistent data sources where the final series of the estimates is derived by pooling together information from these sources.

1.1 Purpose of technical note

This technical note outlines processes undertaken to develop the ward statistical product titled "Ward level Small Area Population Estimates, 2022".

2. Ward product development processes

Preparations of the ward statistical product involved a number of processes including:

- Assessment of Census 2022 data for all geography areas below the local municipality;
- Alignment of Census 2022 data to 2020 ward boundaries;
- Ward population estimation; and
- Preparation of data for publication.

2.1. Assessment of data below local municipality

Data assessment and evaluation exercise was undertaken to ascertain data instability at lower levels. The data assessment exercise was implemented as part of a strategy to identify instances of data instability at the sub-municipal level. This initiative is intended to inform the dissemination strategies for the final phase of the Census 2022 products. An assessment and data confrontation process included identification of critical indicators and an analysis of Census 2022 against Census 2011 data to check patterns, levels and trends for place name layers (main-place and sub-place) and electoral ward layer.

2.1.2 Ward layer data assessment

A two-step process was used in assessing Census 2022 data namely: alignment of census datasets to latest ward boundaries and computation of indicators for person level counts.

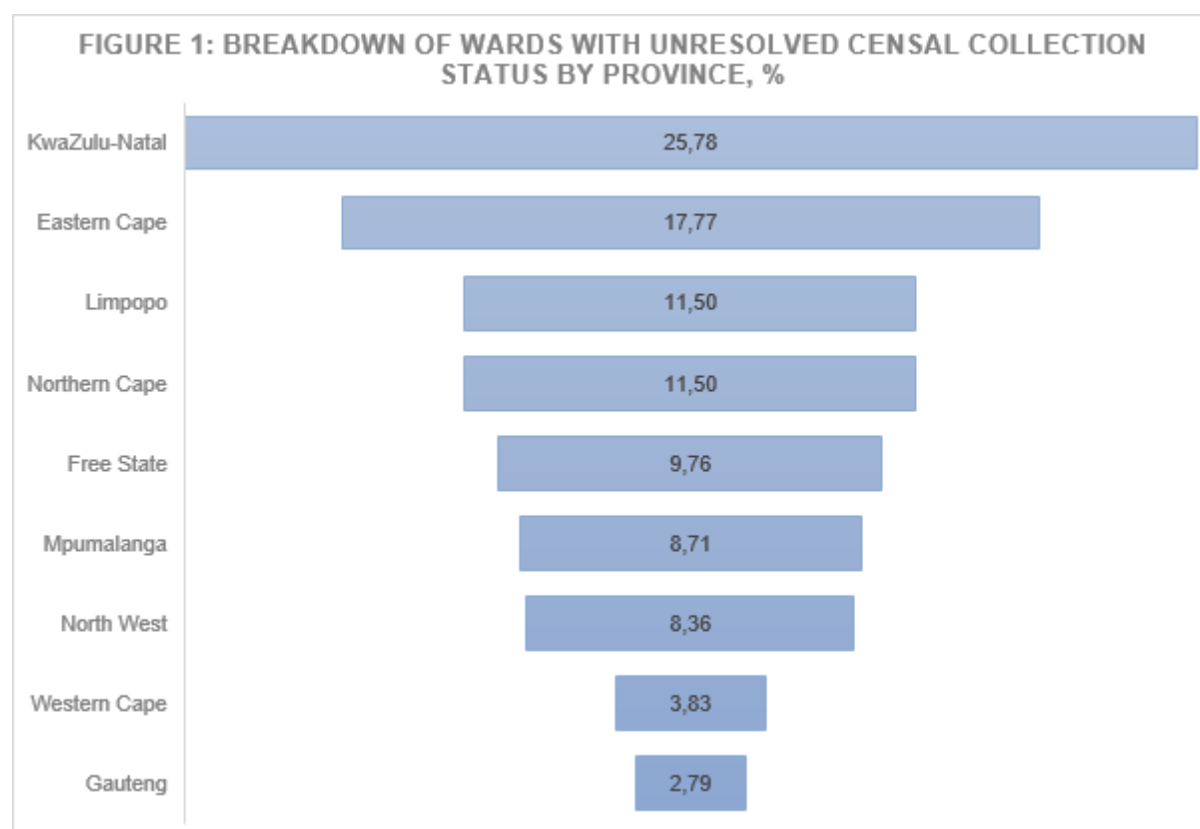**2.1.3 Alignment of Census 2022 data to 2020 wards**

In preparing the ward layer, all 133 364 Enumeration Areas (EAs) in Census 2022 were dropped to a "point" layer using a centroid method linking them to their respective 2020 ward boundaries as applicable to the 4 468 wards identified by the Municipal Demarcation Board. Using a centroid method ensures that all EAs belong to only one ward and not multiple wards. A spatial join was done combining attributes from one class to another based on a specified spatial relationship between the Census 2022 EAs and wards. The spatial join involves matching rows in the ward boundary layer to rows in the Census 2022 EA layer, based on their spatial location. The output of that process is a conjoined layer containing the attributes from both the EA and ward layer.

The procedure of linking what is stated above was also followed to link the 103 576 EAs in Census 2011 with the 2020 ward boundaries for comparison purposes.

1

2.1.4 Ward layer indicator assessment

An assessment of ward data was done which included a number of indicators associated with the population structure at ward level such as population size, age, sex, and population group. The outcome of this assessment is that out of a total of 4 468 wards, 287 wards (6,42%) had an unresolved censal collection status in Census 2022. These wards experienced a negative population growth rate between Census 2011 and Census 2022. Unresolved censal collection refers to cases whose enumeration status cannot be definitively determined.

Figure 1 below present the relative spread of these 287 wards across the country's provinces.

FIGURE 1: BREAKDOWN OF WARDS WITH UNRESOLVED CENSAL COLLECTION STATUS BY PROVINCE, %

Province	%
KwaZulu-Natal	25,78
Eastern Cape	17,77
Limpopo	11,50
Northern Cape	11,50
Free State	9,76
Mpumalanga	8,71
North West	8,36
Western Cape	3,83
Gauteng	2,79

In comparison, when looking at population size, Gauteng ranks as the most populous province (1st), followed by KwaZulu-Natal (2nd), Western Cape (3rd), Eastern Cape (4th), Limpopo (5th), Mpumalanga (6th), North West (7th), Free State (8th), with Northern Cape being the least populous (9th). This ordering is not mirrored in the observed concentration in Figure 1, which suggests that a province's population size is not a determinant of a ward's censal collection status. It follows from this that the occurrence of the identified wards is without a systematic pattern indicating that their unresolved censal collection status is not due to any specific factors.

What follows next is a presentation of the methodology used to derive small area estimates at ward level for South Africa's population pyramid. In particular, Section 2 presents the approach taken, Section 3 reports on the main findings from the approach, and Section 4 provides a conclusion. Any references used and/or consulted are cited in the References section.

2.2 Small area estimation methodology at ward level

Since the auxiliary data for the small area estimation in the present case is confined to two sources, raking, also known as iterative proportional fitting, was employed to generate the

small area estimates for South Africa's population pyramid in 2022 at the ward level. Raking is empirical in nature making no assumptions about the source data from which the small area estimates are derived nor relying on such for the estimation itself. In particular raking is used here to derive small area estimates of the population pyramid at ward level for Age group, Sex, and Population group as they are the variables defining the population pyramid. The population pyramid represents the breakdown of population size by age and gender at a given point in time within and across population groups. Put simply, the population pyramid gives a summary account of the structure of the population at a point in time.

In raking, ratios are computed to scale the rows and columns of a data table to obtain aggregates, i.e. row sums and column sums which are equal to stipulated control totals. In this case, the control totals are from Census 2022, while their initial/known distribution is from Census 2011. The resultant data cells, i.e. the values in the table are adjusted proportionally by these ratios so that the row totals equal the values of the control totals as obtained by summing across the rows. This is followed by re-computation of these ratios such that this time the values in the table are adjusted by them proportionally in order that the column totals equal the values of the control totals as obtained by summing across the columns. This same procedure is successively repeated, i.e. iterated over and over, until the row and column totals are sufficiently close to the control total values used. The procedure is terminated once convergence is reached in the sense that any further iterations produce the same or similar table values.

Raking usually involves two to three passes through the data to achieve convergence with the control totals used. The converged, i.e. iterated estimates are the small area estimates, in this case of the population pyramid in 2022, at ward level per municipality as obtained after two passes.

The first pass involves the following:

- Raking on Age group, Sex, and Population group using the PES-adjusted Census 2022 control totals at the municipality level, on cells populated with Census 2011 population data overlaid with the Census 2022 geography layer for purposes of keeping the same Census 2022 wards between the two points in time.

- Running pairwise comparisons per category for Age group, Sex, and Population group in the 4 181 wards with resolved censal collection status to compare their small area estimates of person level counts for 2022 with the person level counts found in Census 2022, excluding the remaining 287 wards owing to their indeterminate collection status as this affects the reliability of these comparisons for them. To clarify, population size refers to the total number of persons residing within a defined geographical area, in this case a ward. Accordingly, population size and person or person level counts are used interchangeably in this technical note.

The requirement for raking is that the variables; Age group, Sex, and Population group should have the same response categories in both Census 2011 and Census 2022. The Census 2022 data contained the "Unspecified" category for Population group, as well as the "Unspecified" category for Age, while there were no such corresponding categories in the Census 2011 data. To resolve this misalignment in the two data sets, the following adjustments were made to the Census 2011 and Census 2022 data before raking:

- The "Unspecified" Population group was collapsed with the "Other" Population group;
- The count of "Unspecified" Ages was proportionally allocated to the known Ages within each ward; and

3

- Age groups were standardised to 4-yearly intervals, e.g. 0 – 4, 5 – 9, 10 – 14, …, 75 – 79, 80 – 84, and 85+.

Adjustments above, are made for the purpose of calculating control totals at municipal level by Age group, Sex, and Population group such that the response categories in Census 2011 should match the ones in Census 2022. Because it is computationally intensive, a macro is used to rake, with the following specifications:

- The Census 2011 and Census 2022 data sets are aligned in terms of geographic boundaries;

- The municipal control totals are those for population size from Census 2022 for Age group, Sex, and Population group, for purposes of agreement with the published, i.e. reported municipal totals for these variables; and

- The data cell entries are from Census 2011 at the person level comprised of the cross-tabulations between Age group, Sex, and Population group per ward, totalling 416 604 in number across a total of 4 468 wards within a total of 213 municipalities.

After the first pass a second raking pass is done as the derived totals of population size from the first pass disagree with the control totals from the Census 2022 population figures at municipal, provincial and national level. This discrepancy is traced to 58,19% or 167 of the 287 wards with unresolved censal collection status. Consequentially, in the second pass the small area estimates (SAEs) from the first pass for this subset of wards are re-raked with the same Census 2022 control totals for the subset as coming from the first pass.

The outcome of the second raking pass is to preserve the Census 2022 totals for population size at the above-mentioned three levels for all the three variables in the analysis. Finally, the data from the two passes are collated to form the 2022 series of small area estimates for the population pyramid according to its structure by Age group, Sex, and Population group.

3. Main findings

The small area estimates on the population pyramid for 2022 at ward level are obtained by updating the Census 2011 distribution of population size by Age group, Sex, and Population group with the intercensal growth rates between Census 2011 and Census 2022 across these three variables simultaneously. In the current case, these growth rates are represented by the raking ratios. To test the robustness of the raking-derived SAEs, correlation analysis is done using the Pearson correlations between the SAEs and the Census 2022 figures for the 4 181 wards with resolved censal collection status by Age group, Sex, and Population group.

This analysis indicates whether there is equivalence between these two data sets as identified by the strength of their correlations. In this technical note, equivalence has the standard statistical meaning that the values of any two data sets for the same quantity of interest are to be considered practically the same if they are similar enough in general. Table 1 below, summarises the results from the correlation analysis for each of the demographic variables defining the population structure (1A-Sex; 1B-Age Group; and 1C-Population Group).

4

Table 1A: Correlation analysis for person level counts by Sex

Correlation coefficient interval	No. of wards	Cumulative Percent
0,90 – 1,00	4 181	100,00

Table 1B: Correlation analysis for person level counts by Age Group

Correlation coefficient interval	No. of wards	Cumulative Percent
0,90 – 1,00	3 887	92,97
0,70 – 0,89	263	99,26
0,50 – 0,69	20	99,74
0,30 – 0,49	8	99,93
0,00 – 0,29	3	100,00
TOTAL	**4 181**	

Table 1C: Correlation analysis for person level counts by Population Group

Correlation coefficient interval	No. of wards	Cumulative Percent
0,90 – 1,00	4 087	97,75
0,70 – 0,89	76	99,57
0,50 – 0,69	10	99,81
0,30 – 0,49	5	99,93
0,00 – 0,29	1	99,95
-	**2**	100,00
TOTAL	**4 181**	

Note: "-" no correlation computed, as only one population group category is present.

Of the 4 181 wards with resolved censal collection status, more than 99% show very high correlation levels ranging from 0,70 to 1,00 on the unit interval as to their SAEs for population size and this from Census 2022 across any of the demographic variables considered. This empirically confirms equivalence between these two sets, showing that the SAEs of population size are effectively duplicates of the Census 2022 population size for these wards. By extension, the small area estimates of population size by Age group, Sex, and Population group in these 4 181 wards are copied over from Census 2022. In this technical note, information gain is used in its usual statistical sense as the improvement achieved from revising a known distribution to an alternate one. In the current case the known distribution is that from Census 2022 and the alternate one that from the small area estimation.

Turning to the 287 wards with unresolved censal collection status, a closer inspection of the symmetric percentage difference between their SAEs of population size in 2022 and this from Census 2022 shows that raking generates comparatively lower SAEs of population size in 120 or 41,81% of the 287 wards, and conversely higher estimates in 58,19% or 167 wards. Use is made of the symmetric percentage difference, which is the difference between the SAE of population size and its Census 2022 counterpart relative to the average between them expressed in percent. By taking an average, this difference weighs the population size of Census 2011 and Census 2022 equally. This controls against any potential trend bias from choosing either Census as the base of comparison for the intercensal movement of population size between them.

Figure 2 below illustrates the distinctive character of the symmetric percentage difference in the 120, and 167 wards respectively. In the 120 wards, the Census 2022 population size captured by the line of origin exceeds the corresponding raking-derived small area estimates of these wards to a varying extent from ward to ward. In other words, these estimates understate the population size found in Census 2022. With reference to information gain, this indicates that even though convergence to the Census 2022 control totals is not achieved from the first raking pass, replacing the raked estimates of these wards with their Census 2022 person counts would nullify the underestimation that they introduce. As a result, the small area estimates of population size by Age group, Sex, and Population group in these wards are copied over from Census 2022, bringing the number of wards inheriting their Census 2022 data to 4 301 wards.

FIGURE 2: SYMMETRIC PERCENTAGE DIFFERENCES BY WARD NUMBER FOR WARDS WITH UNRESOLVED CENSAL COLLECTION STATUS
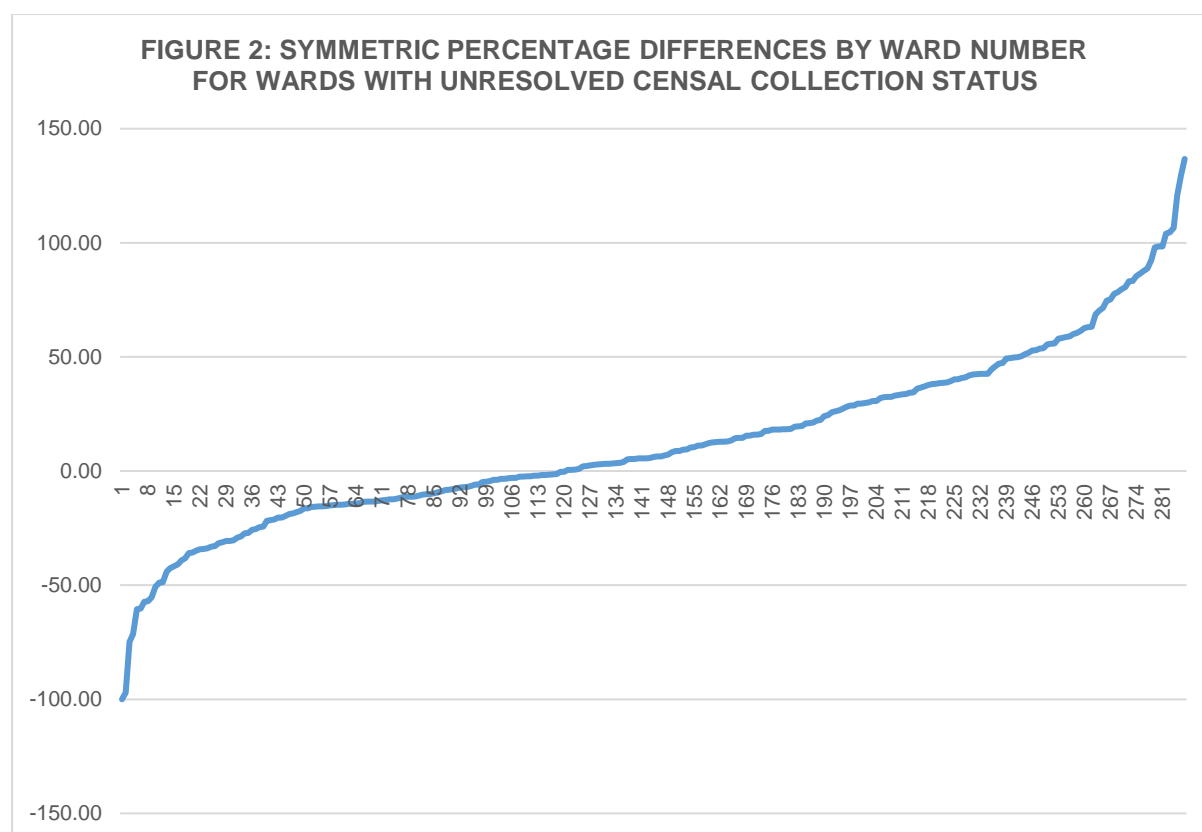
The right-hand side portion of Figure 2 above, also demonstrates that no convergence is achieved by the first raking pass in the remaining 167 wards where the small area estimates of population size are higher from ward to ward compared to the corresponding Census 2022 figures. Here the SAEs in these wards overstate the population size found in Census 2022. This extends to municipal level, and further to provincial and national levels compared to Census 2022. For reasons of compression, the overestimates are summed up by province, showing an overall margin of error of 0,69% nationally, ranging from a low of 0,29% to a high of 1,50% between the provinces. Table 2 below, gives a summary of the incidence of this error margin, computed as the percentage ratio between the "extra" persons added by the first raking pass and the number of persons found in Census 2022.

Table 2: Overestimation of population size from first raking pass, by province

Provinces	No. of wards	"Extra" persons added by first raking pass	Census 2022	Margin of error, %
KwaZulu-Natal	40	67 970	12 423 907	0,55
Eastern Cape	25	54 781	7 230 204	0,76

Free State	20	44 419	2 964 412	1,50
Northern Cape	19	15 299	1 355 945	1,13
Mpumalanga	17	48 528	5 143 324	0,94
Limpopo	16	47 398	6 572 721	0,72
North West	15	57 208	3 804 547	1,50
Western Cape	8	48 990	7 433 020	0,66
Gauteng	7	43 084	15 099 423	0,29
TOTAL	**167**	**427 677**	**62 027 503**	**0,69**

The prevalence of this margin of error is also easy to check, by noting that the incidence of overestimation affects proportionately a higher number of wards (167) compared to that of underestimation (120 wards). To reduce this overestimation of population size introduced by the first raking pass a second raking pass is done, which is isolated to the subset of 167 wards to which it belongs. As per the iterative nature of raking, this second raking pass prorates the small area estimates of population size based on Age group, Sex, and Population group from the first pass of the subset, using its corresponding control totals from Census 2022. This is done while ensuring that the Census 2022 totals for population size remain intact for all 213 municipalities, all 9 provinces, and nationally. Figure 3 below gives a replot of the symmetric percentage differences – this time – for the results from the second raking pass.

As shown in Figure 3 below, after the second raking pass there are 48 wards whose small area estimates of population size coincide with Census 2022, as evidenced by the flat-line segment in the middle of the plot which rests on the line of origin denoting no difference between them. Incidentally this means that there is perfect correlation between these two data sets, indicating that the small area estimates of population size for this segment of wards are duplicates of this found in Census 2022. Consequently, there are an additional 48 wards whose small area estimates of population size by Age group, Sex, and Population group are copied over from Census 2022 since there is no loss in information gain from doing so. This brings the number of wards inheriting their Census 2022 data to 4 349 wards.

Figure 3 also shows how the remaining 119 wards are to be treated with respect to their small area estimates. To the left of the flat-line segment are 65 wards or 54,62% of the 119 wards whose small area estimates of population size understate what was found in Census 2022 with the extent of this varying from ward to ward as seen by the differences in the drop-down lines from the point of origin. To the right of this segment are 54 wards or 45,37% of the 119 wards whose small area estimates of population size overstate what was found in Census 2022. In the latter instance, the extent of this also varies from ward to ward as seen by the differences in the drop-down lines leading to the point of origin.

FIGURE 3:SYMMETRIC PERCENTAGE DIFFERENCES BY WARD NUMBER FOR WARDS INVOLVED IN SECOND RAKING PASS
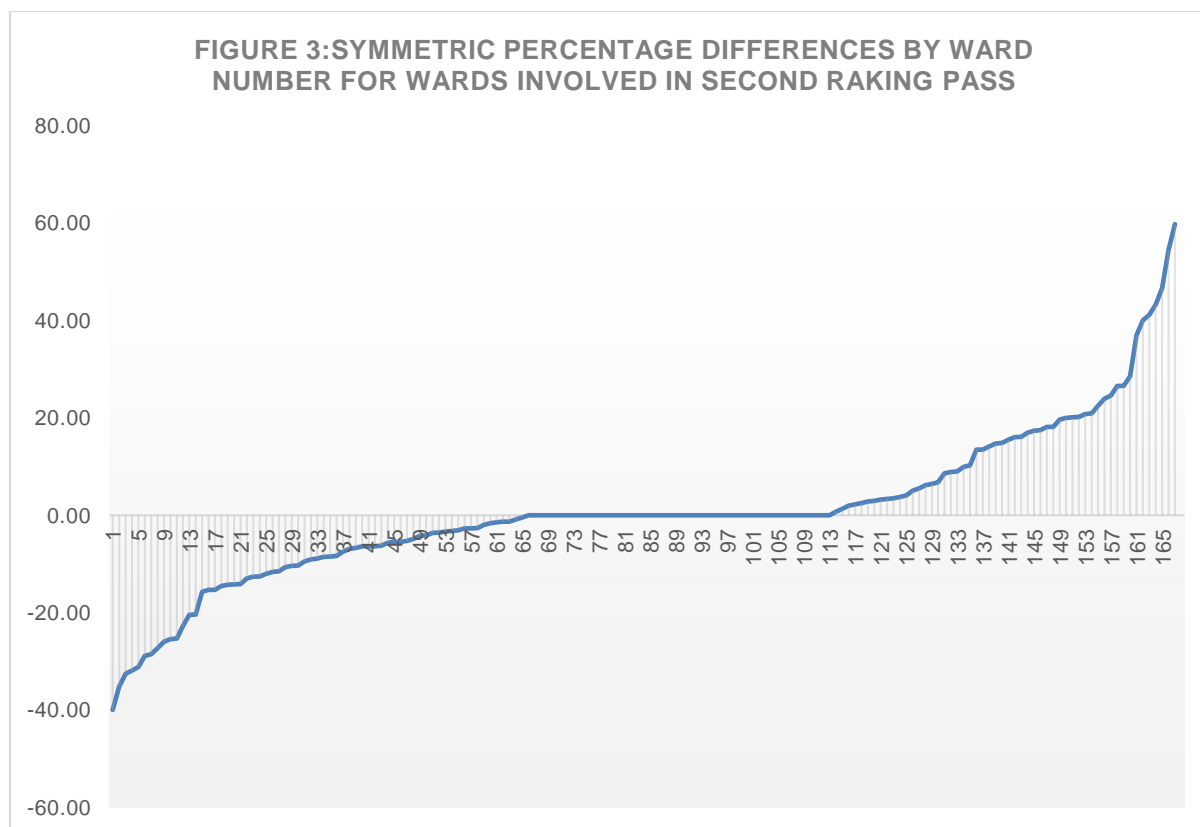
Essentially, the result of the second raking pass is to produce small area estimates for 119 wards whose under and over-estimations tend to cancel each other out on average, as observed in Figure 3, which incidentally is also reflected in the approximate evenness between the left and right portions of the plot. Due to this, when aggregated the SAEs of these wards reconcile with the totals of population size from Census 2022 at the municipal, provincial, and national level in the sense that convergence at these levels is achieved. Compared to the first raking pass information gain is thus realised by the second raking pass since its consolidated small area estimates of population size agree with that found in Census 2022 at these three levels. By extension the raked estimates for these 119 wards from the second raking pass are copied over as their respective small area estimates of population size by Age group, Sex, and Population group in 2022.

The final step of raking is to collate the small area estimates for the population pyramid according to the breakdown of population size in 2022 by Age group, Sex, and Population group. The ensuing joined, i.e. pooled series is comprised of 4 349 wards or 97,34% of the total of 4 468 wards which receive their small area estimates from raking with replacement, by inheriting their Census 2022 data, plus another 119 wards or 2,66% of the total of 4 468 wards who receive their small area estimates from raking without replacement after the latter's second pass. In the former case this is either because the person level counts in Census 2022 are equivalent to those from raking or they control against any misestimation by them. In the latter case these misestimations cancel each other out on average, thus allowing the replacement of the raked estimates to be foregone.

4. Conclusion

This technical note outlines the methodology followed in the derivation of small area estimates (SAEs) of the population pyramid in 2022, at ward level. Of the four different types or classes of estimates, the SAEs derived here are interpolations in terms of which the person level counts from Census 2011 falling in the 4 468 wards of Census 2022 are recursively updated with the latest information regarding their size from the intercensal growth rates between Census 2011 and Census 2022 applicable to the three variables namely Age group, Sex, and

Population group defining the population pyramid. This updating is done by raking whose ratios represent the intercensal growth rates, which when iteratively applied to the person level counts of Census 2011, yield the estimates of population size according to its structure in 2022, at ward level.

It is important to note that for purposes of small area estimation, a geographical area is regarded as small if the area sample needed to survey it is smaller than required to yield precise and reliable estimates about it. The objective of small area estimation is to achieve precision and reliability by instead borrowing strength from existing data sources to arrive at such estimates without surveying either because of the cost implications of requisitioning a survey for each and every area, in this case the ward, or the representativity of the survey sample the area requires would be inadequate to guarantee the validity of its estimates. In the present case, the small area estimation conducted borrows strength from two Censuses namely Census 2011 and Census 2022, the enumeration scale of which ensures that the information required at ward level for this estimation is available. A byproduct of this estimation are the indirect estimates of population size for 2022 effectively obtained as the product of the prorated intercensal growth rates between Census 2011 and Census 2022 applied to the person levels in Census 2011 for Age group, Sex, and Population group. Due to this, in nature the resultant SAEs of population size differ from that of a Census in the sense that under the latter this size is only determined by enumeration at a point in time.

The precision of the SAEs in the current case is verified by the correlations and/or information gain analysis from comparing them to their observed counterparts in Census 2022, while their reliability is confirmed by the convergence they reach under raking – in the sense of achieving a match – with the totals of Census 2022 across the municipal, provincial, and national levels. Following these validations the final SAE series for the population pyramid at Census 2022 ward level is then collated to be harmonised with Census 2022 data on the principle that the latter's distribution is the reference standard to be satisfied.

5. How to access the ward data

The data may be accessed from the Stats SA website on **www.statssa.gov.za**, and is available in a number of formats such as Excel, SuperCross, and SuperWeb.

References

ABS, 2009. *Population estimates: concepts, sources and methods.* Australian Bureau of Statistics: Canberra.

ADB, 2020. *Introduction to small area estimation techniques: a practical guide for national statistics offices.* Asian Development Bank: Manilla.

Chen, H., 2022. Small area estimation for SDGs. *International Workshop on the Monitoring of the Sustainable Development Goals.* United Nations Statistics Division, 12 – 13 January.

Cole, T.J. and Altman, D.G., 2017. What is a percentage difference. *British Medical Journal*, *358*(8), pp. j3663.

Deming, W.E. and Stephan, F.F., 1940. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*, *11*(4), pp.427-444.

Gomatam, S., Karr, A.F. and Sanil, A.P., 2005. Data swapping as a decision problem. *Journal of Official Statistics*, *21*(4), pp.635-655.

Izrael, D., Hoaglin, D.C. and Battaglia, M.P., 2000. A SAS macro for balancing a weighted sample. In *Proceedings of the twenty-fifth annual SAS users' group international conference* (Vol. 275).

Lomax, N. and Norman, P., 2016. Estimating population attribute values in a table: "get me started in" iterative proportional fitting. *Professional Geographer*, *68*(3), pp.451-461.

Purcell, N.J. and Kish, L., 1980. Postcensal estimates for local areas (or domains). *International Statistical Review*, *48*(1), pp.3-18.

Wheeler, J.O., 2005. Geography, in Kempf-Leonard, K. (ed.) *Encyclopedia of Social Measurement*, Elsevier: Amsterdam.