Research Article

An Extension of Karrup–King–Newton Index

Christophe Chesneau,1 Abdul Majeed Lodhi,2 Sadaf Khan,3 Jamal Abdul Nasir,4 Ronald Onyango,5 and Farrukh Jamal3

1Université de Caen Normandie, LMNO, Campus II, Science 3, Caen 14032, France
2Department of Statistics, the Islamia University of Bahawalpur, Rahim Yar Khan Campus, Bahawalpur, Pakistan
3Department of Statistics, the Islamia University of Bahawalpur, Bahawalpur, Pakistan
4Department of Statistics Government College University Lahore, Lahore, Pakistan
5Department of Applied Statistics, Financial Mathematics and Actuarial Science, Jaramogi Oginga Odinga University of Science and Technology, Bondo, Kenya

Correspondence should be addressed to Ronald Onyango; assangaronald@gmail.com

Received 12 October 2021; Accepted 21 January 2022; Published 27 February 2022

Academic Editor: Saima K. Khosa

Copyright © 2022 Christophe Chesneau et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To evaluate the erroneous age reporting, there are several demographic methods available depending upon the type of age data. For example, in the case of single year of age data, some of the most commonly used measures in literature are the Whipple index, Myers index, Bachi index, and modified Whipple indices. This paper proposes a new modification of the Karrup–King–Newton index as a general measure to check the quality of age distortion. When we compare the modified Karrup–King–Newton index to the original index, it appears to be more adequate in terms of the estimation objective. Also, the proposed index is based on the same linearity assumption and practically gives the same result and direction as the original Karrup–King–Newton index.

1. Karrup–King–Newton Index: A Measure of Age Heaping

The Karrup–King–Newton formula is also called Newton’s quadratic formula. It is a well-known fact that age distortion is minimized when the age data are given at a single age and then grouped into the 5-year age interval, but some errors still remain in the 5-year age grouping. Age distortion will be further reduced considerably by the 10-year age group, as errors that transfer beyond the 5-year age group will be captured. This method is based on the following assumptions.

(i) The transfer of age misreporting does not stretch the 5-year age group, so we group into the 10-year age group. This method presumes that the error will be removed.

(ii) The patterns of error are the same in the consecutive six 5-year age groups.

The limitations of this method are the following ones.

(i) The errors across 10-year age groups, most particularly older ages, are not covered.

(ii) They cannot adjust the values of ages 0–9.

(iii) It may smooth out the genuine pattern.

(iv) It makes use of the values of ages 0–9 in the computations, but the error pattern in the age group is different from that of ages 10–70.

The used formulas are given below.

\[ 5^P_x = \frac{1}{210} P_x + \frac{1}{16} (10P_{x-10} - P_{x+10}), \] (1)
and
\[ sP_{x+5} = 10P_x - sP_x, \]  \( \text{(2)} \)

where \( sP_x \) is the first of two 5-year age groups compared to the 10-year age group \( 10P_x \).

2. Modified Karup–King–Newton Index

In the proposed modification, we use the 3-year age group as compared to the 5-year age group or the 10-year age group. The 3-year age interval is better in the sense that it will reduce errors as compared to the 5-year age group. Age misreporting will be further reduced by using 6-year age groupings, as errors that transfer beyond the 3-year age group will be captured. The index is developed on the principle of the original Karup–King–Newton index. The derivation of the index is given below.

Suppose that the population in 3-year age groups is divided into 6 consecutive groups denoted by \( W_0W_1W_2W_3W_4W_5 \).

The 6-year age group is made up of two consecutive 3-year age groups. Let the population be divided into 6-year age groups denoted by \( V_0V_1V_2 \) and \( V_3V_4V_5 \).

Let the 3-year age group be captured. BT the index is developed on the principle of the original Karup–King–Newton index. The derivation of the index is given below.

\( V_0 = W_0 + W_1 \)
\( V_1 = W_2 + W_3 \)
\( V_2 = W_4 + W_5 \)

The difference table using the cumulated populations is given in Table 1.

Note that
\[ \sigma_2 = \frac{V_0}{28} + \frac{V_1}{2} + \frac{V_2}{28} = 0.0357V_0 + 0.5V_1 - 0.0357V_2. \]  \( \text{(3)} \)

This equation is equal to the Karup–King multiplier. The values in the above table are found using the method described below.

\( \sigma_x \) for age \( x \) and \( x + 6 \) are
\( = P + V_0 - P/6 = V_0/6 \), where 6 is the six-year interval.

\( \sigma_x \) for age \( x + 9 \) and \( x + 6 \) are
\( = P + V_0 + U - (P + V_0)/3 = U/3 \), where 3 is the 3-year interval.

\( \sigma_x \) for age \( x + 12 \) and \( x + 9 \) are
\( = P + V_0 + U - (P + V_0 + U)/3 = V_1 - U/3. \)

\( \sigma_x \) for age \( x + 15 \) and \( x + 12 \) are
\( = V_2/12. \)

For the calculation of \( \sigma_2 \) of first term, we have done the following operation:
\( \sigma_2 = 1/3[U/3 - V_0/6] = 1/3[2U - V_0/6]. \)

For calculation of \( \sigma_2 \) of second term, we have done
\( \sigma_2 = 1/6[V_1 - U/3 - U/3] = 1/6[V_1 - 2U/3]. \)

The same method is applied for next differences. \( W_0W_1 \) for age

\( W_0W_1 \) for age

We will need to find the smoothed values for the 3-year age groups equivalent to \( W_2 \) and \( W_3 \), as the center pair of values.

Let us denote with “\( U \)” the smoothed value of \( W_2 \), so the smoothed value of \( W_3 \) will be \( V_1-U \), since the total of the smoothed value should always be equal to the original population data. Let “\( \sigma \)” be the divided differences of the age groups.

From Table 1, a smooth pattern in the cumulated age distribution is given as
\( V_2 - 8V_1 + 14U = 6V_1 + V_0 - 14U. \)
\( 28U = 6V_1 + V_0 - V_2 + 8V_1 \)
\( 28U = 14V_1 + V_0 - V_2 \) OR \( 28U = V_0 - V_2 + 14V_1 \)
\( = 1/28[V_0 - V_2] + 1/2V_1 \)
\( = 1/28V_0 - 1/28V_2 + 1/2V_1. \)  \( \text{(4)} \)

We split 6-year age group into three-year age group.
\( V_1 - U = V_1 - [1/28(V_0 - V_2) + 1/2(V_1)]. \)
\( V_1 - U = V_1 - 1/28(V_0 - V_2) - 1/2(V_1). \)
\( 0.5(V_1) - 0.0357(V_0 - V_2) \) OR \( 0.0357(V_2) + 0.5(V_1) - 0.0357(V_0). \)

It is exciting to see that (4) and (5) are also the same as the Karup–King oscillatory multipliers for splitting 6-year age groups into 3-year age groups.

2.1. Assumption

(i) The transfer of age (net age misreporting error) does not stretch beyond 3 years, so by grouping into a 6-year age group, the method assumes that the error will be cancelled out.

(ii) The patterns of error are similar in six successive 3-year age groups.

2.2. Limitations

(1) The method cannot be used for calculating adjusted values for people aged 72–74 and above.

(2) It may smooth out a genuine pattern.

(3) It makes use of ages 0–4 and ages 5–9 in the computation, whereas error pattern of age groups 10–70 are different from reported age.

(4) The decision to use this method or other methods of adjustment is largely arbitrary.

3. Data Application

The data from the 1998 population and Housing Census of Pakistan and the 2010 Census of Ghana are used to compare the original and modified indices. We see that the original
Karup–King–Newton method adjusts the age groups of 10–69, while the further modified index adjusts the age data of 0–71. Table 2 shows the smoothed population of Pakistan. It indicates that populations in the following age groups: 10–14, 20–24, 25–29, 30–34, 45–49, 50–54, and 60–64, appear to be over-enumerated, while the other age groups are under-reported. Table 3 indicates the result of the modified Karup–King–Newton index in which age groups 0–2, 9–11, 15–17, 21–23, 27–29, 33–35, 42–44, 51–53, and 63–65 appeared to be under-reported, while the other age groups were over-enumerated. Table 4 presents the smoothed population of the Ghana Census. It indicates that age groups 10–14, 25–29, 35–39, 40–44, 50–54, and 60–64 appear to be over-reported, while the other age groups are under-reported. Table 5 shows the result of the modified Karup–King–Newton index of Ghana census data in which age groups 0–2, 6–8, 12–14, 18–20, 24–26, 30–32, 45–47, 48–50, 54–56, 60–62, 69–71, and 75–77 appear to be over-reported while the other age groups are under-reported.

### 4. Graphical Representation of Both Indices

To access the modified Karup–King–Newton index, we compare it with the original Karup–King–Newton index. The graphs for both indices are shown in Figure 1–3. Both indices show the same pattern. The modified Karup–King–Newton index is moderately easy to calculate, allowing the linearity assumptions.

### 5. Final Remarks

In this paper, we used the Karup–King–Newton and modified Karup–King–Newton indices of age heaping to examine the accuracy of age data from Pakistan Census data 1998 and Ghana Census data 2010 and examined the quality of reported data from the population census 1998. In the literature, Tekpli [1] applied five smoothed techniques to Ghana data to check for age misreporting. Also, in the literature, we see that the main problem in age misreporting is lack of education [2–9].

In previous years, researchers used only the Whipple type index, Bachi index, and the Myers blended index to reduce the error in age distribution; no one used the other methods to overcome the error in age misreporting [10, 11]. Our proposed Karup–King–Newton index will produce...
### Table 4: Reported and smoothed population of Ghana.

| Age   | Reported age | Karup–King–Newton index |
|-------|--------------|--------------------------|
| Total, 0–79 | 24,342,445 |                           |
| Total, 10–69 | 17,250,804 | 17,250,804 |
| 0–4 | 3,405,406 |                           |
| 5–9 | 3,128,952 |                           |
| 10–14 | 2,916,040 | 2,898,062 |
| 15–19 | 2,609,989 | 2,627,967 |
| 20–24 | 2,323,491 | 2,338,415 |
| 25–29 | 2,050,111 | 2,035,187 |
| 30–34 | 1,678,809 | 1,690,678 |
| 35–39 | 1,421,403 | 1,409,534 |
| 40–44 | 1,186,350 | 1,171,188 |
| 45–49 | 938,098 | 953,260 |
| 50–54 | 833,098 | 763,067 |
| 55–59 | 523,695 | 593,726 |
| 60–64 | 475,849 | 434,829 |
| 65–69 | 293,871 | 334,891 |
| 70–74 | 351,330 |                           |
| 75–79 | 205,953 |                           |
| 80+ | 316,378 |                           |

### Table 5: Reported and smoothed population of Ghana.

| Age   | Reported | Modified Karup–King–Newton index |
|-------|----------|----------------------------------|
| 0–2 | 2034713 | 2012773 |
| 3–5 | 2023699 | 2045638 |
| 6–8 | 1910612 | 1846015 |
| 9–11 | 1803493 | 1868089 |
| 12–14 | 1677881 | 1609771 |
| 15–17 | 1576126 | 1644235 |
| 18–20 | 1684040 | 1460963 |
| 21–23 | 1260934 | 1484010 |
| 24–26 | 1341828 | 1211738 |
| 27–29 | 1120663 | 1250752 |
| 30–32 | 1191734 | 1055730 |
| 33–35 | 948102 | 1084105 |
| 36–38 | 786629 | 790226 |
| 39–41 | 807003 | 803405 |
| 42–44 | 553094 | 584897 |
| 45–47 | 643284 | 611480 |
| 48–50 | 630087 | 504536 |
| 51–53 | 381795 | 507345 |
| 54–56 | 400431 | 312129 |
| 57–59 | 239294 | 327595 |
| 60–62 | 357557 | 296808 |
| 63–65 | 242830 | 303578 |
| 66–68 | 139021 | 186906 |
| 69–71 | 244840 | 196954 |
| 72–74 | 136802 | 139244 |
| 75–77 | 152288 | 149845 |
| 78–81 | 148405 |                          |
| Total | 24,658,823 |                          |
| 0–77 | 24288780 | 24288780 |
**Figure 1:** Comparison of Karup–King–Newton index with modified Karup–King–Newton index.

**Figure 2:** Reported and smoothed population of Pakistan.
the same direction of errors as the original Karup–King–Newton index. Its main advantages lie in the ease of calculation and its agreement with the original index.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors’ Contributions

Christophe Chesneau gave the initial idea and proposed the new model. Ronald Onyango investigated properties and carried out calculations. Abdul Majeed Lodhi was responsible for the numerical part of the paper. Jamal Abdul Nasir and Farrukh Jamal mutually wrote and structured the manuscript and also verified and investigated numerical calculations applied to the dataset. Sadaf Khan revised the final version according to the reviewers’ comments and also improved the language and presentation of the final version.

References

[1] G. D. Tekpli, *Evaluation and Adjustment of Age Sex Data of Population and Housing Census of Ghana 2000 and 2010*, University of Ghana, Accra, Ghana, 2010.

[2] B. A’Hearn, J. Baten, and D. Crayen, “Quantifying quantitative literacy: age heaping and the history of human capital,” *The Journal of Economic History*, vol. 69, no. 3, pp. 783–808, 2009.

[3] R. Bachi, “The tendency to round off age returns: measurement and correction,” *Bulletin of the International Statistical Institute*, vol. 33, no. 4, pp. 195–222, 1951.

[4] T. Dahiru and H. G. Dikko, “Digit preference in Nigerian censuses data of 1991 and 2006. Epidemiology,” *Biostatistics and Public Health*, vol. 10, no. 2, 2013.

[5] A. Kidane, “Digit preference in African survey data and their impact on parametric estimates,” in *Proceedings of the African Econometric Society Conference*, Abuja, Nigeria, July 2009.

[6] R. J. Myers, “Accuracy of age reporting in the 1950 United States census,” *Journal of the American Statistical Association*, vol. 49, no. 268, pp. 826–831, 1954.

[7] O. Po and H. Tesaghiorghis, *The Population of Zambia*, CICRED, Paris, France, 1975.

[8] E. G. Stockwell, “Patterns of digit preference and avoidance in the age statistics of some recent national censuses: a test of the Turner hypothesis,” *Eugenics Quarterly*, vol. 13, no. 3, pp. 205–208, 1966.

[9] M. Zakria, *Stochastic Models for Population of Pakistan*, Doctoral Dissertation, Allama Iqbal Open University, Islamabad, 2009.

[10] R. D. Retherford and G. M. Mirza, “Evidence of age exaggeration in demographic estimates for Pakistan,” *Population Studies*, vol. 36, no. 2, pp. 257–270, 1982.

[11] F. Yusuf, “On the extent of digital preference in reporting of ages in Pakistan,” *Pakistan Development Review*, vol. 7, no. 4, pp. 519–532, 1967.