Error in age-sex Data of WB

Saikat Roy
IIPS/MBD/2020-22/28

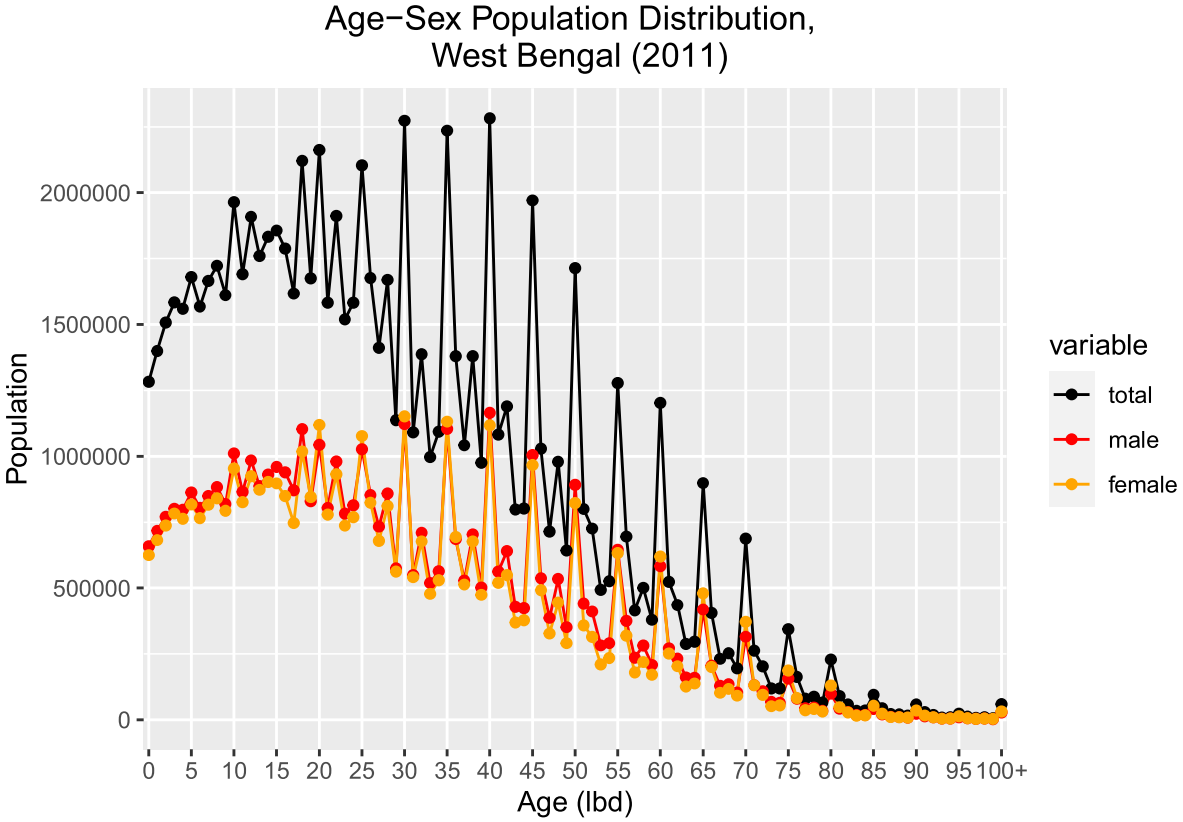
23/12/2020

The errors in the age-sex data of West Bengal as enumerated in the census of West Bengal, 2011 are discussed here.

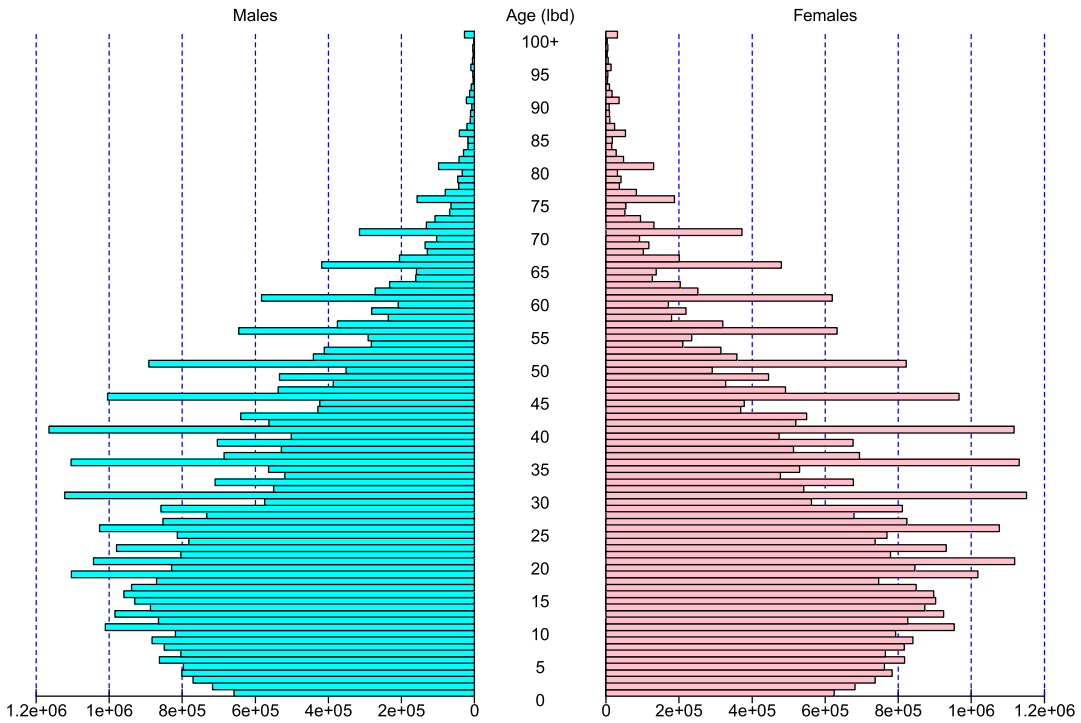
Single-year Age Groups

Graphical Methods

Line curves are plotted for total, male and female populations according as the ages.



Single year population pyramid with absolute population size values for each age-sex is constructed.



It is clear from the above figures that population count unusually spikes hugely at ages ending 0 and 5. Spikes, albeit smaller, are also noticed at even ages.In reality, the age-sex distribution ought to be smooth and devoid of all the spikes. It is clear that people are biased in favour of reporting ages divisible by 5 and slightly lesser biased towards reporting even ages. The tendency to report odd ages is low. This age-heaping is a prime source of errors.

Whipple's Index

The *Whipple's index* measures the extent of preference for ages ending with 0 and 5 under the assumption of rectangularity or linearity assumption in a ten-year age range.

Given that P_x is the population of West Bengal of the cohort with lbd x, the *Whipple's index* for heaping on terminal age digit '0' is measured as,

$$WI_0 = \frac{P_{30} + P_{40} + P_{50} + P_{60}}{\frac{P_{23} + P_{24} + P_{25} + \ldots + P_{62}}{10}} \times 100 \;=\; 161.97$$

and that for heaping on terminal age digit '5' is measured as,

$$WI_5 = \frac{P_{25} + P_{30} + P_{35} + P_{40} + P_{45} + P_{50} + P_{55} + P_{60}}{\frac{P_{23} + P_{24} + P_{25} + \ldots + P_{62}}{5}} \times 100 \;=\; 163.23$$

The *Whipple's indices* for ages ending 0 and 5 are very high. This denotes that the data is not accurate due to high digit preferences for 0 and 5 which results in age-heaping.

Myre's Index

The *Myre's index* for the blended population of West Bengal comes out to be 11.13. The terminal digits, as arranged in descending order are 0, 5, 3, 9, 7, 4, 1, 2, 8, 6. As usual, terminal digits 0 and 5 occupy the highest positions in terminal digit preference. Interestingly, in the case of West Bengal, the terminal digit 3 has high preference. This may also be corroborated from the line diagrams. The population seems to jump at most ages ending with age 3.

Five-Year Age Group

UN Age-Sex Score

The *United Nations age-sex score* gives the inaccuracies in age-sex data of 5 year age-groups.

Let ${}_5M_x$ and ${}_5F_x$ be the male and female population in the x_{th} age-group of a five - year age group population distribution. The *index of sex-ratio score* is given as,

$$SRS = \frac{\sum_{x=5}^{70} \left| \frac{{}_5M_x}{{}_5F_x} - \frac{{}_5M_{x-5}}{{}_5F_{x-5}} \right| \times 100}{n-1}$$

where n is the number of age-groups in the data.

The *index of age-ratio score* is given by,

$$MARS = \frac{\sum_{x=5}^{70} |{}_5MAR_x - 100|}{n-2} \qquad ; \qquad FARS = \frac{\sum_{x=5}^{70} |{}_5FAR_x - 100|}{n-2}$$

where ${}_5MAR_x$ and ${}_5MAR_x$ are the age ratios of males and females at the x_{th} 5 year age group.

The *UN age-sex accuracy index/ joint score* finally comes out to be

$$UNJS = MARS \;+\; FARS \;+ 3 \;\times\; SRS = 21.27$$

This signifies that there are inaccuracies in this 5 year age group data.