A method for calculating the strength and shape of age heaping

Onno Boonstra

In their pioneering work on the Florence Catasto of 1427, Herlihy and Klapish-Zuber (1977), were astonished to see that the ages of the Florentine population were distributed very unevenly. There were much more people with an even numbered age than with an uneven numbered age; there was also distinct age heaping at multiples of 5, 10 and even 12.
Herlihy and Klapish-Zuber did not use statistics to prove the strength and shape of age heaping. More recent research does make use of a few statistical tools, but they are not very refined. A number of indices is being used, for instance Whipple's index, which is meant to detect age heaping at multiples of 5. However, it does not take into consideration age heaping at other multiples. For other multiples, there are other indices who do.
All these indices share a number of problems. First of all, although they do make clear that age heaping exists, they do not measure to what extent age heaping occurs. There are a few rules of the thumb, but these have not been validated statistically. And these rules cannot be compared to one another, so that it cannot become clear which form of age heaping is predominant. Next to that, the fact that there is always a kind of interaction between forms of age heaping (for instance, the heaping at age 60 can be the result of age heaping at multiples of 2, 3, 4, 5, 6, 10, 12, 20 or 30) is not taken into consideration.
A further problem with methods like Whipple's index is that it cannot be used in surveys or with sources where the ages of respondents are limited, like marriage registers.
In my paper, I will introduce a new, simple and straightforward method which (1) calculates the strength of age heaping statistically, (2) while measuring all possible kinds of shapes of age heaping, (3) making the results comparable. Finally, it is a method which (4) can be used in samples with a limited age distribution.