humanleague: a C++ microsynthesis package with R and python interfaces

Andrew P Smith[1]

1 School of Geography and Leeds Institute for Data Analytics, University of Leeds

Summary

humanleague is a microsynthesis package for R and python, with its core implementation in C++. It provides both traditional and novel algorithms for generating synthetic populations from two or more marginal constraints and, optionally, a seed population. The marginal constraints can be of arbitrary dimensionality.

The package provides a fast implementation of the traditional Iterative Proportional Fitting (IPF) algorithm, which generates fractional populations given marginal constraints and a seed population. Where integral populations are preferred, the package also provides two variants of a quasirandom sampling algorithm (QIS) which generate high-entropy 'IPF-like' whole-number populations. The first variant is extremely fast but can only be used where no seed data is provided, and is described in (Smith, Lovelace, and Birkin 2017). The second variant (QIS-I) supports a seed population by sampling from a dynamically-computed IPF population.

The QIS-I algorithm can also be used to integerise precomputed multidimensional fractional populations. Functions are also provided to integerise discrete univariate probability distributions, directly generate quasirandom (Sobol) sequences, and to convert populations represented as counts in a multidimensional state array to a tabular form listing individuals.

References

Smith, Andrew, Robin Lovelace, and Mark Birkin. 2017. "Population Synthesis with Quasirandom Integer Sampling." *Journal of Artificial Societies and Social Simulation* 20 (4):14. https://doi.org/10.18564/jasss.3550.