

Synthetic populations

Why synthetic populations?



- methods comparison (e.g design-based simulation studies)
- policy modelling on individual level (e.g health planning, climate change, demographic change, economic change, . . .)
- teaching (e.g.teaching of survey methods)
- Creation of public-/scientific-use files with low disclosure risk
data availability is often a problem (legal issues, costs, . . .)



- actual sizes of regions and strata need to be reflected marginal distributions and interactions between variables should be represented correctly
- hierarchical and cluster structures have to be preserved Data confidentiality must be ensured
- Pure replication of units from the underlying sample should be avoided Sometimes some marginal distributions must exactly match known values



- choice of methods highly depends on available information survey samples
- aggregated information from samples known marginal distributions from population



- in `simPop`: model-based approach is forced
- Helps to generate synthetic populations, having similar properties as the real population.