

## Some ReGenesees pros w.r.t. the survey package

Feature	How it works
Calibration and variance estimation functions can efficiently process large-scale surveys even in environments with low computational resources (e.g. ordinary PCs)	<ul style="list-style-type: none"> <li>✓ Exploits calibration models factorization through a dedicated divide and conquer algorithm</li> <li>✓ Accelerates execution and saves memory by means of ad-hoc optimizations of many internal functions (e.g. variance, design effects and domain estimation for nonlinear estimators)</li> </ul>
Provides estimates and sampling errors for arbitrary user-defined Complex Estimators, i.e. any nonlinear differentiable function of Horvitz-Thompson or Calibration Estimators of totals	<ul style="list-style-type: none"> <li>✓ Enables users to define their own Complex Estimators symbolically (i.e. as mathematical functions) by means of R expressions</li> <li>✓ Exploits R symbolic differentiation facilities to linearize Complex Estimators automatically, so that their variance is estimated on-the-fly</li> </ul>
Assists users in computing and organizing population totals for calibration tasks	<ul style="list-style-type: none"> <li>✓ Driven by the calibration model formula, automatically generates a template dataframe to be filled with actual population totals</li> <li>✓ If the sampling frame of the survey is available, the template is filled automatically</li> <li>✓ Can cope with sampling frames of several million rows and thousands of auxiliary variables by means of a dedicated adaptive chunking algorithm</li> </ul>
Interaction with all summary statistics functions (i.e. estimators of totals, means, frequencies, ratios, quantiles, multiple regression coefficients, and complex estimators) has been standardized, so that they are easier to assemble in an industrialized process	<ul style="list-style-type: none"> <li>✓ All estimators share (nearly) the same interface, even for domain estimation</li> <li>✓ All estimators produce return values with the same structure, even for subpopulation estimation</li> <li>✓ Estimates and sampling errors can be written to database tables or exported to external files in a common data model</li> </ul>
New statistical capabilities and utilities	<ul style="list-style-type: none"> <li>✓ Hints on feasible bounds for range restricted calibration</li> <li>✓ Quick estimates of auxiliary variables totals</li> <li>✓ Compression of nested factors to reduce model-matrix sparseness in calibration tasks</li> <li>✓ Detailed diagnostics on the calibration process and on its results</li> <li>✓ Merge of new survey data into existing design objects</li> <li>✓ Collapsed strata technique for getting rid of lonely PSUs in variance estimation</li> <li>✓ Detailed diagnostics on the collapsing process and on the generated superstrata</li> <li>✓ Covariance and correlation between Complex Estimators</li> <li>✓ A Generalized Variance Functions (GVF) infrastructure, i.e. facilities for defining, fitting, testing and plotting GVF models, and to exploit them to predict variance estimates</li> </ul>
Provides a comprehensive and user-friendly point-and-click GUI	<ul style="list-style-type: none"> <li>✓ Pure R implementation, relies on tcltk and tcltk2 packages</li> </ul>