ReGenesees

(R evolved Generalised software for sampling estimates and errors in surveys)

Scope

Design-Based and Model-Assisted Analysis of Complex Sample Surveys

Main Statistical Functions

- **Complex Sampling Designs**
  - Multistage, stratified, clustered, sampling designs
  - Sampling with equal or unequal probabilities, with or without replacement
  - "Mixed" sampling designs (i.e. with both Self-Representing and Non-Self-Representing strata)
- **Calibration**
  - Global and partitioned (for factorizable calibration models)
  - Unit-level and cluster-level weights adjustment
  - Homoscedastic and heteroscedastic models
  - Linear, raking and logit distance functions
  - Bounded and unbounded weights adjustment
  - Multi-step calibration
  - Calibration on multiple regression coefficients
  - Consistent trimming of calibration weights
- **Basic Estimators**
  - Horvitz-Thompson
  - Calibration Estimators
- **Variance Estimation**
  - Multistage formulation (via Bellhouse recursive algorithm)
  - Ultimate Cluster approximation
  - Collapsed strata technique for handling lonely PSUs
  - Taylor linearization of nonlinear smooth estimators
  - Generalized Variance Functions (GVF) method
- **Estimates and Sampling Errors (standard error, variance, coefficient of variation, confidence interval, design effect) for:**
  - Totals
  - Means
  - Absolute and relative frequency distributions (marginal, conditional and joint)
  - Ratios between totals
  - Shares and ratios between shares
  - Multiple regression coefficients
  - Quantiles (variance estimation via the Woodruff method)

  - Population variance and standard deviation of numeric variables
  - Measures of Change derived from two not necessarily independent samples
- **Estimates and Sampling Errors for Complex Estimators**
  - Handles arbitrary differentiable functions of Horvitz-Thompson or Calibration estimators
  - Complex Estimators can be freely defined by the user
  - Automated Taylor-linearization
  - Design covariance and correlation between Complex Estimators
- **Estimates and Sampling Errors for Subpopulations (Domains)**
  - All the analyses above can be carried out for arbitrary domains

System Architecture

ReGenesees is a full-fledged software system entirely developed in R. It has a clear-cut two-layer architecture. The application layer of the system is embedded into an R package named itself **ReGenesees**. A second R package, called **ReGenesees.GUI**, implements the presentation layer of the system. Both packages can be run under Windows, Mac, as well as under most of the Unix-like operating systems. While the **ReGenesees.GUI** package requires the **ReGenesees** package, the latter can be used also without the GUI on top. This means that the statistical functions of the system will always be accessible by users interacting with R through the traditional command-line interface. On the contrary, less experienced R users will take advantage from the user-friendly mouse-click graphical interface.

Data Input/Output

The ReGenesees system can import data in a variety of ways. First, it can load R workspace files (with .RData or .rda extensions) storing previously saved data. Second, data can be imported from Text Files (with extensions .txt, .csv, .dat). Third, the system can import data from MS Excel spreadsheets and/or MS Access database tables. Further extensions are possible. Currently, ReGenesees can save output data into R workspace files (.RData, .rda) and/or export them into Text Files (.txt, .csv, .dat). Further extensions are possible.

Development Status

The current version of the ReGenesees system is **2.2**

Software Documentation

Both packages composing the system (**ReGenesees** and **ReGenesees.GUI**) come with their own reference manuals, which fulfill R standards for packages' documentation.

Software Distribution

The ReGenesees system is distributed as Open Source Software, under the EUPL license.

Website

ReGenesees' website is hosted on GITHUB at the following URL:

- https://diegozardetto.github.io/ReGenesees

Authors

Overall Project: Diego Zardetto (zardetto@istat.it)

Application layer (i.e. **ReGenesees** package): Diego Zardetto

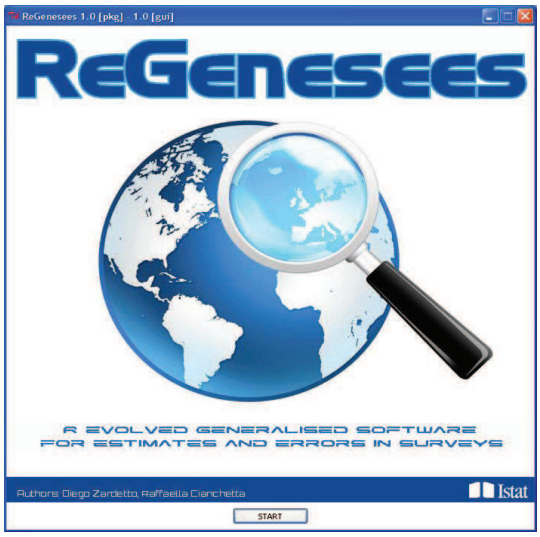
Presentation layer (i.e. **ReGenesees.GUI** package): Diego Zardetto, and Raffaella Cianchetta

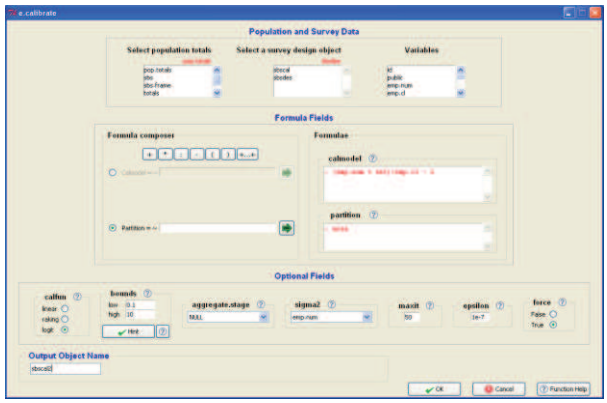
Download

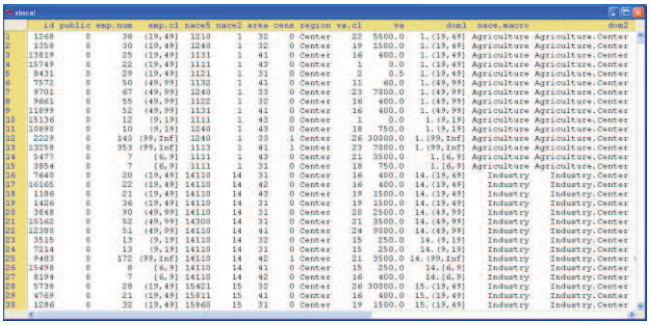
The ReGenesees system can be downloaded from:

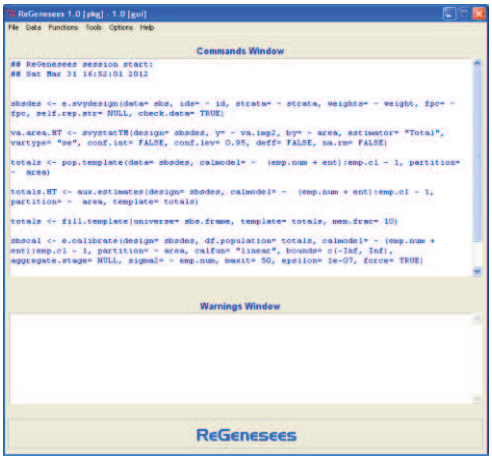
- **Istat website**
  - English:
    http://www.istat.it/en/tools/methods-and-it-tools/processing-tools/regenesees
  - Italian:
    http://www.istat.it/it/strumenti/metodi-e-strumenti-it/strumenti-di-elaborazione/regenesees
- **GITHUB**
  - ReGenesees:
    https://github.com/DiegoZardetto/ReGenesees
  - ReGenesees.GUI:
    https://github.com/DiegoZardetto/ReGenesees.GUI
- **The European Commission Repository for Open Source Software (Joinup):**
  https://joinup.ec.europa.eu/software/regenesees/description

Sample GUI Screenshots









References

- **Woodruff, R. S.** - (1952)
  "Confidence Intervals for Medians and Other Position Measures"
  Journal of the American Statistical Association,
  Vol. 47, n. 260, pp. 635-646.
- **Woodruff, R. S.** - (1971)
  "A Simple Method for Approximating the Variance of a Complicated Estimate"
  Journal of the American Statistical Association,
  Vol. 66, n. 334, pp. 411-414.
- **Wilkinson, G.N., Rogers, C.E.** - (1973)
  "Symbolic Description of Factorial Models for Analysis of Variance"
  Journal of the Royal Statistical Society, series C (Applied Statistics),
  Vol. 22, pp. 181-191.
- **Kalton, G.** - (1979)
  "Ultimate cluster sampling"
  Journal of the Royal Statistical Society, series A (General),
  Vol. 142, Part 2, pp. 210-222.
- **Binder, D. A.** - (1983)
  "On the variances of asymptotically normal estimators from complex surveys"
  International Statistical Review,
  Vol. 51, n. 3, pp. 279-292.
- **Rust, K.** - (1985)
  "Variance Estimation for Complex Estimators in Sample Surveys"
  Journal of Official Statistics,
  Vol. 1, n. 4, pp. 381-397.
- **Bellhouse, DR.** - (1985)
  "Computing Methods for Variance Estimation in Complex Surveys"
  Journal of Official Statistics,
  Vol.1, n. 3, pp. 323-329.
- **Rust, K., Kalton, G.** - (1987)
  "Strategies for Collapsing Strata for Variance Estimation"
  Journal of Official Statistics,
  Vol. 3, n. 1, pp. 69-81.
- **Korn, E.L., Graubard, B.I.** - (1990)
  "Simultaneous testing of regression coefficients with complex survey data: Use of Bonferroni t statistics"
  The American Statistician,
  Vol. 44, n. 4, pp. 270-276.
- **Särndal, C.E., Swensson, B., Wretman, J.** - (1992)
  "Model Assisted Survey Sampling"
  Springer Verlag.
- **Deville, J.C., Särndal, C.E.** - (1992)
  "Calibration Estimators in Survey Sampling"
  Journal of the American Statistical Association,
  Vol. 87, n. 418, pp. 376-382.
- **Chambers, J.M., Hastie, T.J.** - (1992)
  "Statistical Models in S"
  Wadsworth & Brooks/Cole.

- **Deville, J.C., Särndal, C.E., Sautory, O.** - (1993)
  "Generalized Raking Procedures in Survey Sampling"
  Journal of the American Statistical Association,
  Vol. 88, n. 423, pp.1013-1020.
- **Sautory, O.** - (1993)
  "La macro CALMAR: Redressement d'un Echantillon par Calage sur Marges"
  Document de travail de la Direction des Statistiques Demographiques et Sociales,
  n. F9310.
- **Dorfman, A., Valliant, R.** - (1993)
  "Quantile variance estimators in complex surveys"
  Proceedings of the ASA Survey Research Methods Section,
  pp. 866-871.
- **Kish, L.** - (1995)
  "Methods for design effects"
  Journal of Official Statistics,
  Vol. 11, n. 1, pp. 55-77.
- **Estevao, V., Hidiroglou, M. A., Särndal, C. E** - (1995)
  "Methodological principles for a generalized estimation system at Statistics Canada"
  Journal of Official Statistics,
  11, n. 2, pp. 181-204.
- **Singh, A.C., Mohl, C.A.** - (1996)
  "Understanding calibration estimators in survey sampling"
  Survey Methodology,
  22, pp. 107-115.
- **Rao, J. N. K., Lohr, S. L.** - (1999)
  "Some Current Trends in Sample Survey Theory and Methods"
  Sankhya: The Indian Journal of Statistics, Special issue on Sample Surverys,
  Vol. 61, Series B, Pt. 1, pp. 1-57.
- **Valliant, R.** - (2000)
  "Variance estimation for the general regression estimator"
  Survey Methodology,
  28, pp. 103-114.
- **Vanderhoeft, C.** - (2001)
  "Generalized Calibration at Statistic Belgium"
  Statistics Belgium Working Paper n. 3
  http://statbel.fgov.be/nl/binaries/paper03%5B1%5D_tcm325-35412.pdf
- **Fuller, W.A.** - (2002)
  "Regression estimation for survey samples"
  Survey Methodology,
  28, pp. 5-23.
- **Rao, J. N. K., Lohr, S. L.** - (2004)
  "Sample Survey Methods: Recent Developments and Applications"
  two-day workshop slides, Joint Statistical Meetings, Toronto.
- **Lumley, T.** - (2006)
  "survey: analysis of complex survey samples"
  R package version 3.6-5.
  http://cran.at.r-project.org/web/packages/survey/index.html
- **Wolter, K. M.** - (2007)
  "Introduction to Variance Estimation"
  Second Edition, Springer-Verlag, New York.

- **Scannapieco, M., Zardetto, D., Barcaroli, G.** - (2007)
  "La Calibrazione dei Dati con R: una Sperimentazione sull'Indagine Forze di Lavoro ed un Confronto con GENESEES/SAS"
  Contributi Istat n. 4.
  http://www3.istat.it/dati/pubbsci/contributi/Contributi/contr_2007/2007_4.pdf
- **Lumley, T.** - (2012)
  "*Complex Surveys: A Guide to Analysis Using R*"
  John Wiley & Sons, New York.
- **Barcaroli, G., Zardetto, D.** - (2012)
  "Use of R in Business Surveys at the Italian National Institute of Statistics: Experiences and Perspectives"
  Proceedings of the 4th International Conference of Establishment Surveys (ICES IV), American Statistical Association.
  http://www.amstat.org/meetings/ices/2012/papers/302193.pdf
- **Zardetto, D.** - (2013)
  "ReGenesees: an Advanced R System for Calibration, Estimation and Sampling Errors Assessment in Complex Sample Surveys"
  Proceedings of the 7th International Conference on New Techniques and Technologies for Statistics (NTTS 2013), Eurostat.
  http://www.cros-portal.eu/sites/default/files//NTTS2013fullPaper_131-v2.pdf
- **Fallows A., Pope M., Digby-North J., Brown G., Lewis D.** - (2015)
  "A Comparative Study of Complex Survey Estimation Software in ONS"
  Romanian Statistical Review,
  n. 3, pp. 46-64.
  http://www.revistadestatistica.ro/index.php/comparative-study-of-complex-survey-estimation-software-in-ons/
- **Zardetto, D.** - (2015)
  "ReGenesees: an Advanced R System for Calibration, Estimation and Sampling Error Assessment in Complex Sample Surveys"
  Journal of Official Statistics,
  Vol. 31, n. 2, pp. 177-203.
  https://sciendo.com/article/10.1515/jos-2015-0013