

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
 - Unequally weighted sampling, with or without replacement
 - “Mixed” sampling designs (i.e. with both SelfRepresenting and NonSelfRepresenting strata)
- **Calibration**
 - Global and/or partitioned (for factorizable calibration models)
 - Unit-level and/or cluster-level adjustment
 - Homoscedastic and/or Heteroscedastic models
- **Basic Estimators**
 - Horvitz-Thompson
 - Calibration Estimators
- **Sampling Variance Estimation**
 - Multistage formulation (via Bellhouse recursive algorithm)
 - Ultimate-Cluster approximation
 - Taylor-linearization for nonlinear “smooth” estimators
 - Collapse strata technique for handling lonely PSUs
- **Estimates and Sampling Errors (standard errors, variance, coefficient of variation, confidence interval, design effect) for:**
 - Totals
 - Means
 - Absolute and/or relative frequency distributions (marginal, conditional and joint)
 - Ratios between totals
 - Multiple regression coefficients
 - Quantiles (variance estimation via the Woodruff method)
- **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)**

Under development:

- Generalized Variance Functions (GVF) method.
- Replication based Variance Estimation for non-analytic estimators, through the Delete-A-group Jackknife (DAGJK) technique: this will integrate the [EVER](#) package with the ReGenesees system.

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 its 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 **1.4.0**

Software Documentation

Both packages composing the system (**ReGenesees** and **ReGenesees.GUI**) come with their own reference manuals, which fulfill R standards for packages' documentation. In addition, a more extended and self-contained user guide for the whole system is currently under development.

Software Distribution

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

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, Raffaella Cianchetta

Download

The ReGenesees system can be downloaded from:

- The European Commission Repository for Open Source Software:
<http://joinup.ec.europa.eu/software/regenesees/release/all>
- Istat website (web pages in Italian):
<http://www.istat.it/it/strumenti/metodi-e-software/software/regenesees>

Sample GUI Screenshots

The image displays three screenshots of the ReGenesees GUI. The top-left screenshot shows the main window with the 'ReGenesees' logo and a globe icon. The top-right screenshot shows the 'e.calibrate' dialog box, which is used for selecting population totals, survey design objects, and variables, and for setting formula fields and optional fields. The bottom screenshot shows the 'Commands Window' with a list of commands and their results, including the 'e.calibrate' command and its output.

ReGenesees 1.0 [pkg] - 1.0 [gui]

ReGenesees

R EVOLVED GENERALISED SOFTWARE
FOR ESTIMATES AND ERRORS IN SURVEYS

Humans: Diego Zardetto, Raffaella Cianchetta

Istat

START

e.calibrate

Population and Survey Data

Select population totals: pop.totals, pop.frame, totals

Select a survey design object: shadef, shades

Variables: id, public, emp.num, emp.ci

Formula Fields

Formula composer: Calculated, Partition

Formulas: calmodel, partition

Optional Fields: bounds, lower, upper, log, aggregate.stage, sigma2, maxit, epsilon, force

Output Object Name: shadef

ReGenesees 1.0 [pkg] - 1.0 [gui]

File Data Functions Tools Options Help

Commands Window

```
## ReGenesees session start:
## Sat Mar 31 16:52:01 2012

shades <- e.svydesign(data= shd, id= id, strata= strata, weights= weight, fpc= fpc, seif.rep.st= NULL, check.data= TRUE)

va.area.HT <- svyestatT(design= shades, y= va.imp2, by= area, estimator= "Total",
vartype= "se", conf.int= FALSE, conf.level= 0.95, deff= FALSE, na.rm= TRUE)

totals <- pop.template(data= shades, calmodel= (emp.num + ent) * emp.ci - 1, partition= area)

totals.HT <- svyestatT(design= shades, calmodel= (emp.num + ent) * emp.ci - 1, partition= area)

totals <- fill.template(universe= shd.frame, template= totals, mem.trac= 10)

shadef <- e.calibrate(design= shades, df.population= totals, calmodel= (emp.num + ent) * emp.ci - 1, partition= area, calfun= "linear", bounds= c(-Inf, Inf),
aggregate.stage= NULL, sigma2= emp.num, maxit= 50, epsilon= 1e-07, force= TRUE)
```

Warnings Window

ReGenesees

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,
51, pp. 279-292.
- **Wolter, K. M.** - (1985)
"Introduction to Variance Estimation"
Springer-Verlag, New York.
- **Rust, K.** - (1985)
"Variance Estimation for Complex Estimators in Sample Surveys"
Journal of Official Statistics,
1, pp. 381-397.
- **Bellhouse, DR.** - (1985)
"Computing Methods for Variance Estimation in Complex Surveys"
Journal of Official Statistics,
Vol.1, No.3, pp. 323-329.
- **Rust, K., Kalton, G.** - (1987)
"Strategies for Collapsing Strata for Variance Estimation"
Journal of Official Statistics,
Vol. 3, No. 1, pp. 69-81.
- **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, No. 418, pp. 376-382.
- **Deville, J.C., Särndal, C.E., Sautory, O.** - (1993)
"Generalized Raking Procedures in Survey Sampling"
Journal of the American Statistical Association,
Vol. 88, No. 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,
 no. 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, pp. 55-77.
- **Estevao, V., Hidirolou, 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 Surveys,
 Volume 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://www.statbel.fgov.be/studies/paper03_en.asp.
- **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>
- **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://www.istat.it/dati/pubbsci/contributi/Contributi/contr_2007/2007_4.pdf