<http://my.ilstu.edu/~mxu2/Spring-2020/MAT450/MTH450.html>

## survey: Analysis of Complex Survey Samples

Summary statistics, two-sample tests, rank tests, generalised linear models, cumulative link models, Cox models, loglinear models, and general maximum pseudolikelihood estimation for multistage stratified, cluster-sampled, unequally weighted survey samples. Variances by Taylor series linearisation or replicate weights. Post-stratification, calibration, and raking. Two-phase subsampling designs. Graphics. PPS sampling without replacement. Principal components, factor analysis.

|  |  |
| --- | --- |
| Version: | 3.37 |
| Depends: | R (≥ 3.1.0), grid, methods, [Matrix](http://cran.fhcrc.org/web/packages/Matrix/index.html), [survival](http://cran.fhcrc.org/web/packages/survival/index.html) |
| Imports: | stats, graphics, splines, [lattice](http://cran.fhcrc.org/web/packages/lattice/index.html), [minqa](http://cran.fhcrc.org/web/packages/minqa/index.html), [numDeriv](http://cran.fhcrc.org/web/packages/numDeriv/index.html), [mitools](http://cran.fhcrc.org/web/packages/mitools/index.html) (≥ 2.4) |
| Suggests: | [foreign](http://cran.fhcrc.org/web/packages/foreign/index.html), [MASS](http://cran.fhcrc.org/web/packages/MASS/index.html), [KernSmooth](http://cran.fhcrc.org/web/packages/KernSmooth/index.html), [hexbin](http://cran.fhcrc.org/web/packages/hexbin/index.html), [RSQLite](http://cran.fhcrc.org/web/packages/RSQLite/index.html), [quantreg](http://cran.fhcrc.org/web/packages/quantreg/index.html), parallel, [CompQuadForm](http://cran.fhcrc.org/web/packages/CompQuadForm/index.html), [DBI](http://cran.fhcrc.org/web/packages/DBI/index.html), [AER](http://cran.fhcrc.org/web/packages/AER/index.html) |
| Published: | 2020-01-21 |
| Author: | Thomas Lumley |
| Maintainer: | "Thomas Lumley" <t.lumley at auckland.ac.nz> |
| License: | [GPL-2](http://cran.fhcrc.org/web/licenses/GPL-2) | [GPL-3](http://cran.fhcrc.org/web/licenses/GPL-3) |
| URL: | <http://r-survey.r-forge.r-project.org/survey/> |
| NeedsCompilation: | no |
| Citation: | [survey citation info](http://cran.fhcrc.org/web/packages/survey/citation.html) |
| Materials: | [NEWS](http://cran.fhcrc.org/web/packages/survey/NEWS) |
| In views: | [OfficialStatistics](http://cran.fhcrc.org/web/views/OfficialStatistics.html), [SocialSciences](http://cran.fhcrc.org/web/views/SocialSciences.html), [Survival](http://cran.fhcrc.org/web/views/Survival.html) |
| CRAN checks: | [survey results](http://cran.fhcrc.org/web/checks/check_results_survey.html) |

#### Downloads:

|  |  |
| --- | --- |
| Reference manual: | [survey.pdf](http://cran.fhcrc.org/web/packages/survey/survey.pdf) |
| Vignettes: | [Estimates in subpopulations](http://cran.fhcrc.org/web/packages/survey/vignettes/domain.pdf) [Two-phase designs in epidemiology](http://cran.fhcrc.org/web/packages/survey/vignettes/epi.pdf) [Obsolete formulas for two-phase variances](http://cran.fhcrc.org/web/packages/survey/vignettes/phase1.pdf) [Analysing PPS designs](http://cran.fhcrc.org/web/packages/survey/vignettes/pps.pdf) [A survey analysis example](http://cran.fhcrc.org/web/packages/survey/vignettes/survey.pdf) |
| Package source: | [survey\_3.37.tar.gz](http://cran.fhcrc.org/src/contrib/survey_3.37.tar.gz) |
| Windows binaries: | r-devel: [survey\_3.37.zip](http://cran.fhcrc.org/bin/windows/contrib/4.0/survey_3.37.zip), r-devel-gcc8: [survey\_3.36.zip](http://cran.fhcrc.org/bin/windows/contrib/r-devel-gcc8/survey_3.36.zip), r-release: [survey\_3.37.zip](http://cran.fhcrc.org/bin/windows/contrib/3.6/survey_3.37.zip), r-oldrel: [survey\_3.37.zip](http://cran.fhcrc.org/bin/windows/contrib/3.5/survey_3.37.zip) |
| OS X binaries: | r-release: [survey\_3.37.tgz](http://cran.fhcrc.org/bin/macosx/el-capitan/contrib/3.6/survey_3.37.tgz), r-oldrel: [survey\_3.37.tgz](http://cran.fhcrc.org/bin/macosx/el-capitan/contrib/3.5/survey_3.37.tgz) |
| Old sources: | [survey archive](https://CRAN.R-project.org/src/contrib/Archive/survey) |

#### Reverse dependencies:

|  |  |
| --- | --- |
| Reverse depends: | [CalibrateSSB](http://cran.fhcrc.org/web/packages/CalibrateSSB/index.html), [cjoint](http://cran.fhcrc.org/web/packages/cjoint/index.html), [hopit](http://cran.fhcrc.org/web/packages/hopit/index.html), [lavaan.survey](http://cran.fhcrc.org/web/packages/lavaan.survey/index.html), [mapStats](http://cran.fhcrc.org/web/packages/mapStats/index.html), [MedSurvey](http://cran.fhcrc.org/web/packages/MedSurvey/index.html), [MOJOV](http://cran.fhcrc.org/web/packages/MOJOV/index.html), [pedgene](http://cran.fhcrc.org/web/packages/pedgene/index.html), [relaimpo](http://cran.fhcrc.org/web/packages/relaimpo/index.html), [sae2](http://cran.fhcrc.org/web/packages/sae2/index.html), [samplingbook](http://cran.fhcrc.org/web/packages/samplingbook/index.html), [sptm](http://cran.fhcrc.org/web/packages/sptm/index.html), [ssfit](http://cran.fhcrc.org/web/packages/ssfit/index.html), [StatMatch](http://cran.fhcrc.org/web/packages/StatMatch/index.html), [svydiags](http://cran.fhcrc.org/web/packages/svydiags/index.html), [svyPVpack](http://cran.fhcrc.org/web/packages/svyPVpack/index.html), [twang](http://cran.fhcrc.org/web/packages/twang/index.html), [VIMGUI](http://cran.fhcrc.org/web/packages/VIMGUI/index.html), [weightTAPSPACK](http://cran.fhcrc.org/web/packages/weightTAPSPACK/index.html) |
| Reverse imports: | [aGE](http://cran.fhcrc.org/web/packages/aGE/index.html), [anthro](http://cran.fhcrc.org/web/packages/anthro/index.html), [capm](http://cran.fhcrc.org/web/packages/capm/index.html), [causaldrf](http://cran.fhcrc.org/web/packages/causaldrf/index.html), [convey](http://cran.fhcrc.org/web/packages/convey/index.html), [cregg](http://cran.fhcrc.org/web/packages/cregg/index.html), [dgo](http://cran.fhcrc.org/web/packages/dgo/index.html), [DHS.rates](http://cran.fhcrc.org/web/packages/DHS.rates/index.html), [dvmisc](http://cran.fhcrc.org/web/packages/dvmisc/index.html), [EffectLiteR](http://cran.fhcrc.org/web/packages/EffectLiteR/index.html), [effects](http://cran.fhcrc.org/web/packages/effects/index.html), [egor](http://cran.fhcrc.org/web/packages/egor/index.html), [GB2](http://cran.fhcrc.org/web/packages/GB2/index.html), [GJRM](http://cran.fhcrc.org/web/packages/GJRM/index.html), [httk](http://cran.fhcrc.org/web/packages/httk/index.html), [ICS](http://cran.fhcrc.org/web/packages/ICS/index.html), [ICtest](http://cran.fhcrc.org/web/packages/ICtest/index.html), [jskm](http://cran.fhcrc.org/web/packages/jskm/index.html), [jsmodule](http://cran.fhcrc.org/web/packages/jsmodule/index.html), [jstable](http://cran.fhcrc.org/web/packages/jstable/index.html), [mase](http://cran.fhcrc.org/web/packages/mase/index.html), [MatchThem](http://cran.fhcrc.org/web/packages/MatchThem/index.html), [microsynth](http://cran.fhcrc.org/web/packages/microsynth/index.html), [OmnibusFisher](http://cran.fhcrc.org/web/packages/OmnibusFisher/index.html), [paramhetero](http://cran.fhcrc.org/web/packages/paramhetero/index.html), [Plasmode](http://cran.fhcrc.org/web/packages/Plasmode/index.html), [PNADcIBGE](http://cran.fhcrc.org/web/packages/PNADcIBGE/index.html), [poliscidata](http://cran.fhcrc.org/web/packages/poliscidata/index.html), [pricesensitivitymeter](http://cran.fhcrc.org/web/packages/pricesensitivitymeter/index.html), [rareGE](http://cran.fhcrc.org/web/packages/rareGE/index.html), [RNHANES](http://cran.fhcrc.org/web/packages/RNHANES/index.html), [robsurvey](http://cran.fhcrc.org/web/packages/robsurvey/index.html), [srvyr](http://cran.fhcrc.org/web/packages/srvyr/index.html), [StroupGLMM](http://cran.fhcrc.org/web/packages/StroupGLMM/index.html), [SUMMER](http://cran.fhcrc.org/web/packages/SUMMER/index.html), [SvyNom](http://cran.fhcrc.org/web/packages/SvyNom/index.html), [tab](http://cran.fhcrc.org/web/packages/tab/index.html), [tableone](http://cran.fhcrc.org/web/packages/tableone/index.html), [Zelig](http://cran.fhcrc.org/web/packages/Zelig/index.html) |
| Reverse suggests: | [apyramid](http://cran.fhcrc.org/web/packages/apyramid/index.html), [BIFIEsurvey](http://cran.fhcrc.org/web/packages/BIFIEsurvey/index.html), [broom](http://cran.fhcrc.org/web/packages/broom/index.html), [car](http://cran.fhcrc.org/web/packages/car/index.html), [finalfit](http://cran.fhcrc.org/web/packages/finalfit/index.html), [ggeffects](http://cran.fhcrc.org/web/packages/ggeffects/index.html), [grattan](http://cran.fhcrc.org/web/packages/grattan/index.html), [hutils](http://cran.fhcrc.org/web/packages/hutils/index.html), [inca](http://cran.fhcrc.org/web/packages/inca/index.html), [inctools](http://cran.fhcrc.org/web/packages/inctools/index.html), [insight](http://cran.fhcrc.org/web/packages/insight/index.html), [interactions](http://cran.fhcrc.org/web/packages/interactions/index.html), [ipw](http://cran.fhcrc.org/web/packages/ipw/index.html), [jtools](http://cran.fhcrc.org/web/packages/jtools/index.html), [logmult](http://cran.fhcrc.org/web/packages/logmult/index.html), [parameters](http://cran.fhcrc.org/web/packages/parameters/index.html), [PracTools](http://cran.fhcrc.org/web/packages/PracTools/index.html), [Qtools](http://cran.fhcrc.org/web/packages/Qtools/index.html), [rdhs](http://cran.fhcrc.org/web/packages/rdhs/index.html), [RDS](http://cran.fhcrc.org/web/packages/RDS/index.html), [SDaA](http://cran.fhcrc.org/web/packages/SDaA/index.html), [sirt](http://cran.fhcrc.org/web/packages/sirt/index.html), [sjPlot](http://cran.fhcrc.org/web/packages/sjPlot/index.html), [sjstats](http://cran.fhcrc.org/web/packages/sjstats/index.html), [WeightIt](http://cran.fhcrc.org/web/packages/WeightIt/index.html) |
| Reverse enhances: | [margins](http://cran.fhcrc.org/web/packages/margins/index.html), [prediction](http://cran.fhcrc.org/web/packages/prediction/index.html), [stargazer](http://cran.fhcrc.org/web/packages/stargazer/index.html), [texreg](http://cran.fhcrc.org/web/packages/texreg/index.html) |

#### Linking:

Please use the canonical form <https://CRAN.R-project.org/package=survey> to link to this page.

# Survey analysis in R

This is the homepage for the ["survey"](http://cran.fhcrc.org/web/packages/survey/index.html) package, which provides facilities in [R](http://www.r-project.org) for analyzing data from complex surveys. The current version is 3.29. A much earlier version (2.2) was published in [Journal of Statistical Software](http://www.jstatsoft.org/v09)

An experimental package for very large surveys such as the American Community Survey can be found [here](http://sqlsurvey.r-forge.r-project.org)

A port of a much older version of the survey package (version 3.6-8) to S-PLUS 8.0 is available from [CSAN](http://csan.insightful.com) (thanks to Patrick Aboyoun at Insightful).

Features:

* Means, totals, ratios, quantiles, contingency tables, regression models, loglinear models, survival curves,rank tests, for the whole sample and for domains.
* Variances by Taylor linearization or by replicate weights (BRR, jackknife, bootstrap, multistage bootstrap, or user-supplied)
* Multistage sampling with or without replacement.
* PPS sampling with or without replacement: Horvitz-Thompson and Yates-Grundy estimators and a range of approximations.
* Post-stratification, generalized raking/calibration, GREG estimation, trimming of weights.
* Two-phase designs. Estimated weights for augmented IPW estimators.
* Graphics
* Support for using multiply imputed data
* Database-backed design objects for large data sets (now with replicate weights, too)
* Some support for parallel processing on multicore computers.
* Multivariate analysis: principal components, factor analysis (experimental).
* Likelihood ratio (Rao-Scott) tests for glms, Cox models, loglinear models.

The [NEWS](http://r-survey.r-forge.r-project.org/survey/NEWS) file gives a history of features and bug fixes.

**Comparison shopping:**  
Alan Zaslavsky keeps a comprehensive [list of survey analysis software](http://www.hcp.med.harvard.edu/statistics/survey-soft/) for the ASA Section on Survey Research Methods.

User-generated ratings and reviews of this package (and others) at [crantastic.](http://crantastic.org/packages/survey)

Using the survey package:

* [Specifying a survey design](http://r-survey.r-forge.r-project.org/survey/example-design.html)
* [Creating replicate weights](http://r-survey.r-forge.r-project.org/survey/example-svrepdesign.html)
* [Simple summary statistics](http://r-survey.r-forge.r-project.org/survey/example-summary.html)
* [Using supplied replicate weights](http://r-survey.r-forge.r-project.org/survey/example-svrepdesign1.html)
* [Domain (subpopulation) estimation](http://r-survey.r-forge.r-project.org/survey/example-domain.html)
* [Tables of summary statistics](http://r-survey.r-forge.r-project.org/survey/example-table.html)
* [Post-stratification and calibration](http://r-survey.r-forge.r-project.org/survey/example-poststrat.html)
* [Lonely PSUs](http://r-survey.r-forge.r-project.org/survey/exmample-lonely.html)
* [Regression models](http://r-survey.r-forge.r-project.org/survey/example-regression.html)
* [Tests of association](http://r-survey.r-forge.r-project.org/survey/example-chisq.html)
* [Stratification within PSUs](http://r-survey.r-forge.r-project.org/survey/example-twostage.html)
* [Graphics](http://r-survey.r-forge.r-project.org/survey/example-graphics.html)
* [Multiple imputation and ordinal logistic regression](http://r-survey.r-forge.r-project.org/survey/svymi.html)
* [Database-backed survey objects](http://r-survey.r-forge.r-project.org/survey/svy-dbi.html)
* [Programming with survey objects](http://r-survey.r-forge.r-project.org/survey/example-programming.html)

**Technical notes and comparisons with other software**

Some [examples](http://r-survey.r-forge.r-project.org/survey/ucla-examples.pdf) (in PDF) translated from Stata and SUDAAN examples at [UCLA Academic Technology Services](http://www.ats.ucla.edu/stat/survey/survey_howtochoose.htm).

Notes on the [sparse matrix algorithms](http://r-survey.r-forge.r-project.org/survey/ht-sparse.pdf) used in version 3.15 for two-phase designs (and perhaps more widely in future versions)

Notes on [standard errors for survival curves](http://r-survey.r-forge.r-project.org/survey/survcurve.pdf).

A 2009 [CDC report](http://www.cdc.gov/HealthyYouth/YRBS/pdf/YRBS_analysis_software.pdf) compared five **other** survey analysis packages in the context of the Youth Risk Behaviors Survey. I have written an [extension](http://r-survey.r-forge.r-project.org/survey/YRBS-report-extension.pdf) that does the same feature comparisons and results comparisons with R and the survey package. Some of this is copied from the CDC report (which I believe is in the public domain), but they are (of course) not responsible for any of the conclusions or results.

Anthony Damico has R scripts for downloading and analysing major US government surveys at [Github](https://github.com/ajdamico/usgsd). He reported on [comparisons of the survey package with SAS, Stata, SUDAAN](http://journal.r-project.org/archive/2009-2/RJournal_2009-2_Damico.pdf) in [The R Journal](http://journal.r-project.org) 1(2) 37-45

**Tutorials**

I have [a course at statistics.com](http://www.statistics.com/survey-r), which will be repeated as demand permits

Here are [slides](http://faculty.washington.edu/tlumley/survey-jsm-nup.pdf) from a Continuing Education course at JSM 2012.

I gave a [workshop on two-phase designs](http://www.epiresearch.org/congress/wkshpdesc.php?i=0018) at the [3rd North American Congress of Epidemiology](http://www.epicongress2011.org), in Montreal, June 21,2011

I gave a two-day course for the [Washington (DC) Statistical Society](http://www.scs.gmu.edu/~wss/shortcs.html), March 23-24 2010. First day on [R](http://r-survey.r-forge.r-project.org/survey/R-wss.pdf), second day on the [survey package](http://r-survey.r-forge.r-project.org/survey/survey-wss-2010.pdf)

Norman Breslow and I gave the course at [STATISTICALPS 2009](http://www.statmed.medicina.unimib.it/statisticalps2009/statisticalps.htm), at the beginning of September in the Italian Alps. The course will include an introduction to the survey package, but will focus on two-phase designs in epidemiology. We will have some code and data up soon.

[Slides](http://r-survey.r-forge.r-project.org/survey/survey-census.pdf) from a short tutorial at the US Census Bureau, August 10.

I gave a [tutorial](http://r-survey.r-forge.r-project.org/tutorials/index.html) at [useR 2009](http://www.agrocampus-ouest.fr/math/useR-2009/), on the afternoon of July 7, 2009.

A 1.5 hour [brief introduction to R](http://r-survey.r-forge.r-project.org/survey/aapoR.pdf), including a bit on the survey package, at the AAPOR conference, Friday May 15, 10:30am.

There was a [one-day course](http://staff.pubhealth.ku.dk/~pd/survey-2009) at the University of Copenhagen Center for Health and Society on April 3, 2009. Slides are available at that link.

Tobias Verbeke has [packaged data sets and exercises](http://cran.fhcrc.org/web/packages/SDaA/index.html) from Sharon Lohr's Sampling: Design and Analysis for use with the survey package.

I gave a short course for the Washington Statistical Society on March 15-16 2007. The first day was on R and the slides are a selection from [these](http://faculty.washington.edu/tlumley/Rcourse/). The second day was on the survey package, [slides here](http://r-survey.r-forge.r-project.org/survey/survey-wss.pdf).

Norman Breslow and I gave a short course on complex survey designs for epidemiology at the 2008 WNAR (Biometric Society) meeting, [UC Davis, June 22, 2008](http://www.wnar.org). My sessions were [an overview](http://r-survey.r-forge.r-project.org/survey/survey-wnar.pdf) of the survey package and an introduction to [calibration](http://r-survey.r-forge.r-project.org/survey/wnar-calibrate.pdf). Norm's data sets and code are also [online](http://faculty.washington.edu/norm/WNAR08.html)

There is an article on version 3.6-12 of the package in the [January 2008 issue](http://isi.cbs.nl/iass/N57.pdf) of Survey Statistician (note: large PDF file)

[[book 
cover 
picture](http://r-survey.r-forge.r-project.org/svybook/)](http://r-survey.r-forge.r-project.org/svybook/)I have written a book on survey analysis, based around the survey package. The book is called Complex Surveys: a guide to analysis using R. It has just been published by [John Wiley & Sons.](http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470284307,descCd-description.html) It already has a [web site](http://r-survey.r-forge.r-project.org/svybook/)

**Help pages:**

* [Vignettes](http://r-survey.r-forge.r-project.org/survey/doc/index.html)
* PDF from [CRAN](http://cran.us.r-project.org/src/contrib/Descriptions/survey.html)
* [HTML](http://search.r-project.org/library/survey/html/00Index.html)

The [PEAS](http://www.napier.ac.uk/depts/fhls/peas/) project at Napier University has Practical Examplars for the Analysis of Surveys using R (as well as other packages).

[Thomas Lumley](mailto:tlumley@u.washington.edu)