CRAN Task View: Robust Statistical Methods

Maintainer:	Martin Maechler
Contact:	Martin.Maechler at R-project.org
Version:	2021-11-23
URL:	https://CRAN.R-project.org/view=Robust

Robust (or "resistant") methods for statistics modelling have been available in S from the very beginning in the 1980s; and then in R in package `stats`. Examples are `median()`, `mean(*, trim =. )`, `mad()`, `IQR()`, or also `fivenum()`, the statistic behind `boxplot()` in package `graphics`) or `lowess()` (and `loess()`) for robust nonparametric regression, which had been complemented by `runmed()` in 2003. Much further important functionality has been made available in recommended (and hence present in all R versions) package MASS (by Bill Venables and Brian Ripley, see *the* book Modern Applied Statistics with S). Most importantly, they provide `rlm()` for robust regression and `cov.rob()` for robust multivariate scatter and covariance.

This task view is about R add-on packages providing newer or faster, more efficient algorithms and notably for (robustification of) new models.

Please send suggestions for additions and extensions to the task view maintainer .

An international group of scientists working in the field of robust statistics has made efforts (since October 2005) to coordinate several of the scattered developments and make the important ones available through a set of R packages complementing each other. These should build on a basic package with "Essentials", coined robustbase with (potentially many) other packages building on top and extending the essential functionality to particular models or applications. Since 2020 and the 2nd edition of *Robust Statistics: Theory and Methods* , RobStatTM covers its estimators and examples, notably by importing from robustbase and rrcov. Further, there is the quite comprehensive package robust, a version of the robust library of S-PLUS, as an R package now GPLicensed thanks to Insightful and Kjell Konis. Originally, there has been much overlap between 'robustbase' and 'robust', now robust *depends* on robustbase and rrcov, where 'robust' provides convenient routines for the casual user while robustbase and rrcov contain the underlying functionality, and provide the more advanced statistician with a large range of options for robust modeling.

We structure the packages roughly into the following topics, and typically will first mention functionality in packages robustbase, rrcov and robust.

Regression

- ***Linear*** *Regression:*
  `lmrob()` (robustbase) and `lmRob()` (robust) where the former uses the latest of the fast-S algorithms and heteroscedasticity and autocorrelation corrected (HAC) standard errors, the latter makes use of the M-S algorithm of Maronna and Yohai (2000), automatically when there are factors among the predictors (where S-estimators (and hence MM-estimators) based on resampling typically badly fail). The `ltsReg()` and `lmrob.S()` functions are available in robustbase, but rather for comparison purposes. `rlm()` from MASS had been the first widely available implementation for robust linear models, and also one of the very first MM-estimation implementations. robustreg provides very

simple M-estimates for linear regression (in pure R). Note that Koenker's quantile regression package [quantreg](#) contains L1 (aka LAD, least absolute deviations)-regression as a special case, doing so also for nonparametric regression via splines. Package [mblm](#) 's function `mblm()` fits median-based (Theil-Sen or Siegel's repeated) simple linear models.

- ***Generalized** Linear Models (**GLM** s) for Regression:*
GLMs are provided both via `glmrob()` ([robustbase](#)) and `glmRob()` ([robust](#)). Robust ordinal regression is provided by [rorutadis](#) (UTADIS). [drgee](#) fits "Doubly Robust" Generalized Estimating Equations (GEEs), [complmrob](#) does robust linear regression with compositional data as covariates. [multinomRob](#) fits overdispersed multinomial regression models for count data.
- ***Mixed-Effects** (Linear and Nonlinear) Regression:*
Quantile regression (and hence L1 or LAD) for mixed effect models, is available in package [lqmm](#). Rank-based mixed effect fitting from package [rlme](#), whereas an *MM-like* approach for robust linear **mixed effects** modeling is available from package [robustlmm](#). More recently, [skewlmm](#) provides robust linear mixed-effects models **LMM** via scale mixtures of skew-normal distributions.
- ***Nonlinear / Smooth** (Nonparametric Function) Regression:*
Robust Nonlinear model fitting is available through [robustbase](#) 's `nlrob()`. [robustgam](#) fits robust GAMs, i.e., robust Generalized Additive Models.

Multivariate Analysis:

- Here, the [rrcov](#) package which builds (" `Depends` ") on [robustbase](#) provides nice S4 class based methods, more methods for robust multivariate variance-covariance estimation, and adds robust PCA methodology.
- 'rrcov' is extended by [rrcovNA](#), providing robust multivariate methods for *for incomplete* or missing (`NA`) data, and by [rrcovHD](#), providing robust multivariate methods for *High Dimensional* data.
- Specialized robust PCA packages are [pcaPP](#) (via Projection Pursuit), [rpca](#) (incl "sparse") and [rospca](#). Historically, note that robust PCA can be performed by using standard R's `princomp()`, e.g., `X <- stackloss; pc.rob <- princomp(X, covmat= MASS::cov.rob(X))`
- Here, [robustbase](#) contains a slightly more flexible version, `covMcd()` than [robust](#) 's `fastmcd()`, and similarly for `covOGK()`. OTOH, [robust](#) 's `covRob()` has automatically chosen methods, notably `pairwiseQC()` for large dimensionality p. Package [robustX](#) for experimental, or other not yet established procedures, contains `BACON()` and `covNCC()`, the latter providing the neighbor variance estimation (NNVE) method of Wang and Raftery (2002), also available (slightly less optimized) in [covRobust](#).
- [RobRSVD](#) provides a robust Regularized Singular Value Decomposition.
- [mvoutlier](#) (building on [robustbase](#)) provides several methods for outlier identification in high dimensions.
- [GSE](#) estimates multivariate location and scatter in the presence of missing data.
- [RSKC](#) provides **R** obust **S** parse **K** -means **C** lustering.
- [robustDA](#) for *robust mixture Discriminant Analysis* (RMDA) builds a mixture model classifier with noisy class labels.
- [robcor](#) computes robust pairwise correlations based on scale estimates, particularly on `FastQn()`.
- [covRobust](#) provides the nearest neighbor variance estimation (NNVE) method of Wang and Raftery (2002).

Clustering (Multivariate):

- We are *not* considering cluster-resistant variance (/standard error) estimation (aka "sandwich"). Rather e.g. model based and hierarchical clustering methodology with a particular emphasis on robustness: Note that [cluster](#) 's `pam()` implementing "partioning around medians" is partly robust (medians instead of very unrobust k-means) but is *not* good enough, as e.g., the k clusters could consist of k-1 outliers one cluster for the bulk of the remaining data.
- "Truly" robust clustering is provided by packages [genie](#), [Gmedian](#), [otrimle](#) (trimmed MLE model-based) and notably [tclust](#) (robust trimmed clustering).
- See also the CRAN task views [Multivariate](#) and [Cluster](#)

Large Data Sets:

- `BACON()` (in [robustX](#)) should be applicable for larger (n,p) than traditional robust covariance based outlier detectors.
- [OutlierDM](#) detects outliers for replicated high-throughput data. (See also the CRAN task view [MachineLearning](#).)

Descriptive Statistics / Exploratory Data Analysis:

- `boxplot.stats()`, etc mentioned above

Time Series:

- R's `runmed()` provides *most robust* running median filtering.
- Package [robfilter](#) contains robust regression and filtering methods for univariate time series, typically based on repeated (weighted) median regressions.
- The [RobPer](#) provides several methods for robust periodogram estimation, notably for irregularly spaced time series.
- Peter Ruckdeschel has started to lead an effort for a robust time-series package, see [robust-ts](#) on R-Forge.
- Further, robKalman, "*Routines for Robust Kalman Filtering --- the ACM- and rLS-filter*" , is being developed, see [robkalman](#) on R-Forge.

Econometric Models:

- Econometricians tend to like HAC (heteroscedasticity and autocorrelation corrected) standard errors. For a broad class of models, these are provided by package [sandwich](#); similarly [clubSandwich](#) and [clusterSEs](#). Note that `vcov(lmrob())` also uses a version of HAC standard errors for its robustly estimated linear models. See also the CRAN task view [Econometrics](#)

Robust Methods for Bioinformatics:

- There are several packages in the [Bioconductor project](#) providing specialized robust methods. In addition, [RobLoxBioC](#) provides infinitesimally robust estimators for preprocessing omics data.

Robust Methods for Survival Analysis:

- Package [coxrobust](#) provides robust estimation in the Cox model.

Robust Methods for Surveys:

- On R-forge only, package [rhte](#) provides a robust Horvitz-Thompson estimator.

Geostatistics:

- Package [georob](#) aims at robust geostatistical analysis of spatial data, such as kriging and more.

Collections of **Several** Methodologies:

- [WRS2](#) contains robust tests for ANOVA and ANCOVA and other functionality from Rand Wilcox's collection.
- [walrus](#) builds on [WRS2](#) 's computations, providing a different user interface.
- [robeth](#) contains R functions interfacing to the extensive RobETH fortran library with many functions for regression, multivariate estimation and more.

Other Approaches to robust and resistant methodology:

- The package [distr](#) and its several child packages also allow to explore robust estimation concepts, see e.g., [distr](#) on R-Forge.
- Notably, based on these, the project [robast](#) aims for the implementation of R packages for the computation of optimally robust estimators and tests as well as the necessary infrastructure (mainly S4 classes and methods) and diagnostics; cf. M. Kohl (2005). It includes the R packages [RandVar](#), [RobAStBase](#), [RobLox](#), [RobLoxBioC](#), [RobRex](#). Further, [ROptEst](#), and [ROptRegTS](#).
- [RobustAFT](#) computes Robust Accelerated Failure Time Regression for Gaussian and logWeibull errors.
- [robumeta](#) for robust variance meta-regression; [metaplus](#) adds robustness via t- or mixtures of normal distributions.
- [ssmrob](#) provides robust estimation and inference in sample selection models.

`CRAN packages:`

- [clubSandwich](#)
- [cluster](#)
- [clusterSEs](#)
- [complmrob](#)
- [covRobust](#)
- [coxrobust](#)
- [distr](#)
- [drgee](#)
- [genie](#)
- [georob](#)
- [Gmedian](#)
- [GSE](#)
- [lqmm](#)
- [MASS](#) (core)
- [mblm](#)
- [metaplus](#)
- [multinomRob](#)
- [mvoutlier](#)
- [otrimle](#)

- [OutlierDM](#)
- [pcaPP](#)
- [quantreg](#)
- [RandVar](#)
- [rlme](#)
- [RobAStBase](#)
- [robcor](#)
- [robeth](#)
- [robfilter](#)
- [RobLox](#)
- [RobLoxBioC](#)
- [RobPer](#)
- [RobRex](#)
- [RobRSVD](#)
- [RobStatTM](#)
- [robumeta](#)
- [robust](#) (core)
- [RobustAFT](#)
- [robustbase](#) (core)
- [robustDA](#)
- [robustgam](#)
- [robustlmm](#)
- [robustreg](#)
- [robustX](#)
- [ROptEst](#)
- [ROptRegTS](#)
- [rorutadis](#)
- [rospca](#)
- [rpca](#)
- [rrcov](#) (core)
- [rrcovHD](#)
- [rrcovNA](#)
- [RSKC](#)
- [sandwich](#)
- [skewlmm](#)
- [ssmrob](#)
- [tclust](#)
- [walrus](#)
- [WRS2](#)

Related links:

- [Mailing list: R Special Interest Group on Robust Statistics](#)
- [Robust Statistics in R (TU Vienna)](#)
- R-Forge Project: [distr](#)
- R-Forge Project: [robast](#)
- R-Forge Project: [robkalman](#)
- R-Forge Project: [robust-ts](#)