Change log for R-package **relaimpo**

Version 2.0 (November 22nd, 2007)

- **relaimpo** now works on data frames, linear models and formulae which include factors (treating all dummies for the factor as one group).
- **relaimpo** now allows interactions (currently second-order only), treating these as hierarchically below main effects, i.e. excluding models with interactions included while any of the main effects is not. This feature works with metric **lmg** only. Currently, it cannot be combined with self-chosen groups (that have not been created by **relaimpo** itself for accomodating factors).
- **relaimpo** can now accommodate observation weights. Note that there are different types of observation weights which can all be treated alike for calc.relimp (calculation of metrics only) but have to be treated differently for confidence intervals.
    a. Weights that reflect different variances of the response values, e.g. if response values are estimated coefficients from various studies with different variances, are typically proportional to the inverse variance (the less uncertain the value, the more weight it is given, as in the Aitken estimator for linear models). Such weights can be specified with the new **weights=** option. They are used in calculating the metrics, but each observation is given the same probability in resampling. NOTE that it is inappropriate to use such data with the fixed-x's bootstrap, since the residuals cannot be assumed to be basically exchangeable.
    b. If the weights in a data frame represent the multiplicity of each observation (i.e. there are several units with identical combination of values in the data, and the weights represent the number of units with exactly this value pattern for each row of the data frame), they can also be used in calc.relimp for calculating the metrics. However, such frequency weights cannot be appropriately accomodated in boot.relimp; instead, the data frame with frequencies has to be expanded to include one row for each unit before using the resampling routine (e.g. using function **untable** from package **reshape** or function **expand.table** from package **epitools**; a future version of **relaimpo** may do that for you, currently you have to do it yourself).
    c. If the weights in a data frame represent the number of units of the population that this single observed unit represents, calc.relimp also works correctly, when simple given these weights. However, for obtaining confidence intervals, the data have to be treated as data from a complex survey (which they usually are anyway in such a situation). You have to define the design yourself using package survey. Data from complex surveys can now also be handled by **relaimpo**.
- **relaimpo** now can accommodate data from complex sample surveys. Implementation of this feature heavily relies on package **survey**. Consequently, **relaimpo** now requires the package `survey` to be available.
  The user needs to familiarize her/himself with the package `survey` in order to prepare a design object or a linear model of class **svyglm** to be handed to the functions in package **relaimpo**.
  While calculation of metrics only uses the weights from the design object (which could also be handed to **relaimpo** as a separate weights vector, cf. previous bullet point), estimation of confidence intervals also makes use of the structure from the survey (e.g.

strata, clusters) by making the package `survey` determine the bootstrap weights.
So far, bootstrapping based on such data must be considered **experimental** in the following sense: survey analysts typically use these bootstrap weights for calculating variances only – whether they are under all circumstances usable for e.g. percentile confidence intervals is not absolutely clear. It is considered likely that percentile intervals are a safer option than more advanced intervals like bca intervals.
Note that bootstrapping survey data currently does not work in combination with factors or interactions.

- **`relaimpo`** has a new function **`mianalyze.relimp`** for relative importance assessments based on multiply imputed data sets.
  Note that **`mianalyze.relimp`** only works together with groups, factors, interactions or x-variables calculated on the fly, if calculation of confidence intervals is suppressed (**`no.CI=TRUE`**), so that the function is used as a convenience tool for aggregating calculations from a list of multiply imputed data frames. Confidence intervals currently cannot be obtained in the presence of groups, factors, interactions or calculated x-variables.
  Furthermore, note that **`mianalyze.relimp`** is experimental and approximate in the following sense:
  - It uses asymptotic t-distribution-based confidence intervals based on Rubin's degrees of freedom although the estimated percentages have not been derived als maximum likelihood estimates
  - The intervals thus obtained are forced to be symmetric; since the distributions of percentages are often far from symmetric, the intervals can be approximate only.

  Thus, confidence intervals should be used for rough indication only (even more so than the other bootstrap confidence intervals in **`relaimpo`**, which have also been observed to only approximately satisfy their nominal coverage probabilities.).
- Several users have asked for **`relaimpo`** to cover mixed models. While this issue has so far not been worked on, note that clustered data (e.g. observation of two eyes for each person, observation of several time points for each unit etc.) can be adequately addressed by using the `design=` option for handing a clustered design (defined in package `survey`) to package **`relaimpo`**. (Of course, this is not the same as fitting a mixed model.)
- **`relaimpo`** now includes the call in all computational output objects for proper traceability of what exactly has been done.
- The default for bootstrap confidence intervals has been changed from `bca` to `perc`, since percentile intervals are much faster to compute and have not proven worse than BCa intervals in simulations for this application. Sorry for any inconveniences this change may cause.
- Bug corrected: For booteval.relimp, numbers were truncated instead of rounded to four digits.
- Bug corrected: str() did not work on objects of class bootrelimpeval, because initialization of the object was not formally correct. The functionality of the package was not affected by this bug.

Version 1.2-2 (September 30th 2007)

Changes from previous versions (globally licenced CRAN version without pmvd)

- Corrected description text regarding licencing info, point readers to non-US version on homepage again (was wrong since Version 1.2)

Version 1.2-1 (September 28th 2007)

Changes from previous versions (globally licenced CRAN version without pmvd)

- Eliminated warnings for R 2.6.0 regarding encoding of description file
- Adapted to GPL notation standard

Version 1.2 (January 27th 2007)

Changes from previous versions

- Regressors can now be grouped, which
    - allows to handle large numbers of regressors as long as they are combined into a reasonably small number of groups
    - will in the future allow to handle factors via grouped dummy variables
- Improvements to error messages

Version 1.1-1 (September 21st 2006)

Changes from previous versions

Package gave ERROR on R CMD check for R 2.4.0 alpha, presumably because of changed behavior of automatic printing for S4 objects (error because of empty slots). This has been fixed by creating S4 methods for show and print.

Version 1.1 (June 29th 2006)

Changes from previous versions

Bug fix for the formula method for calc.relimp and boot.relimp: formulae with "." on the right-hand side did cause an error message.

Version 1.0-1 (June 19th 2006)

Changes from previous versions

Global version on CRAN only: Correction to the description file, which for version 1.0 erroneously claimed that this were the non-US version of the package.

Version 1.0 (June 16th 2006)

Changes from previous versions

Several improvements have been made:

- It is now possible to designate some regressors as adjustment regressors that are adjusted out before assessing relative importance of the remaining regressors (option always for functions calc.relimp and boot.relimp).

- Function calc.relimp has been made generic with methods for formula and linear model objects. The default method has also been enhanced to accept more different types of input.
  Overall, the first object handed over to calc.relimp can now be any of the following:
  a covariance matrix (former parameter covg),
  a data matrix or data frame the first column of which needs to be the response variable (like in function lm),
  a response vector (if a regressor matrix x is also provided),
  a linear model formula,
  or a linear model object (class lm).
  Note, however, that **`relaimpo`** does not accept factors as regressors.
- Function boot.relimp has been made generic with methods for formula and linear model objects. The default method has also been enhanced to accept more different types of input. Except for a covariance matrix that is not sufficient for the bootstrapping routine, boot.relimp accepts the same objects as calc.relimp.
- Besides a bootstrapping routine for random regressors, a bootstrapping routine for fixed regressors is now also available (option fixed=TRUE in boot.relimp).
- If data vectors, matrices or frames include missing values,
  `relaimpo` uses complete cases only (based on function complete.cases from package stats) and prints a warning message.
  Options regarding na.action are in effect only if the formula specification of the model is used.
  Previously, a missing value in the data for boot.relimp would have caused an error.
  (Naturally, a covariance matrix given to calc.relimp must not have any missing values.)
- The plots are annotated in a more useful way (overall title, better axis labels, annotation indicating what options were chosen in the calculations).
- Annotation of printed output has been enhanced in line with annotation of plots.
- Two bugs regarding output of booteval.relimp have been fixed:
  - For rank=TRUE and norank=FALSE: If shares or confidence bounds were very small, the printed numbers were far too large (all calculations were correct, but a formatting issue showed cut-off scientific notation).
  - For rank=FALSE or norank=TRUE: the empty line between several metrics showed 0.0000 instead of blanks.

The following changes have been made to settings and defaults (apologies to any pioneering users who may be inconvenienced by one of these)

- **`relaimpo`** no longer works for R-versions before 2.2.1
  (calculations do work from 2.0.1, but number printout can be wrong!)
- The default for option rela has been changed from TRUE to FALSE – sorry for any inconvenience this may cause to pioneering users.
- The default number of bootstrap resamples has been reduced from b=1500 to b=1000.

Version 0.5-1 (April 13th 2006)

This change applies to the non-US version only: PMVD gave an error message, if after leaving out regressors with coefficients estimated as 0 there were less than two remaining regressors. This issue has been fixed.

Version 0.5 (February 3rd 2006)

Change from previous versions

The files $.relimplm.R and $.relimplmbooteval.R have been renamed to dollar.relimplm.R and dollar.relimplmbooteval.R respectively in order to eliminate warnings in checks of R development version.

Version 0.4 (January 21st 2006)

Change from previous versions

Bootstrapping and evaluation of bootstrap results now also works for two regressors only. Previously, this did not work due to three bugs:

- The internal function nchoosek produced a “subscript out of bounds” error (on this occasion, reference to the original package vsn within the function was also corrected; previously, erroneously referenced e1071).
- The internal function last.calc had a bug for two regressors only.
- The function booteval.relimp did not like to receive a numeric value instead of a 1x1 matrix.

Version 0.3 (January 12th 2006)

Change from version 0.2

correction to column labelling of outputs from function calc.relimp:
In version 0.2, column labels of calculated metrics were in the order the metrics were requested and did not fit the calculated metrics, which were in the standard order of the metrics (lmg, pmvd, last, first, betasq, pratt).

Version 0.2 (December 16th 2005)

Changes from version 0.1

1. correction to coerce method for relimplm (as.relimplm.R) and its documentation: coerce method coerces the full object to list, including the non-numeric components
2. R-code tidied, and comments improved/corrected in many files
3. percentages and ranks metrics output by relimplm are named
4. placing of column names for differences output is improved
5. vignette and change log added to documentation

Original version: Version 0.1 (December 1st 2005)

Ulrike Grömping
TFH Berlin – University of Applied Sciences
http://www.tfh-berlin.de/~groemp/