Package ‘rptR’

January 4, 2018

Title  Repeatability Estimation for Gaussian and Non-Gaussian Data
Version  0.9.21
Depends  R (>= 3.2.1)
Date  2018-01-04
Author  Martin Stoffel <martin.adam.stoffel@gmail.com>,
        Shinichi Nakagawa <s.nakagawa@unsw.edu.au>,
        Holger Schielzeth <holger.schielzeth@uni-jena.de>
Maintainer  Martin Stoffel <martin.adam.stoffel@gmail.com>
Description  Estimating repeatability (intra-class correlation) from Gaussian, binary, proportion and count data.
License  GPL (>= 2)
Imports  methods, stats, lme4, parallel (>= 3.1.2), pbapply
Suggests  testthat, knitr, rmarkdown
RoxygenNote  6.0.1
VignetteBuilder  knitr
NeedsCompilation  no
Repository  CRAN
Date/Publication  2018-01-04 10:02:32 UTC

R topics documented:

BeetlesBody .............................................................................. 2
BeetlesFemale .......................................................... 2
BeetlesMale ........................................................................ 3
plot.rpt .......................................................... 4
print.rpt ......................................................... 5
print.summary.rpt .................................................... 6
rpt .......................................................... 6
rptBinary .............................................................. 10
rptGaussian ................................................................. 13
rptPoisson ................................................................. 16
BeetlesBody dataset

Details

This is an simulated dataset which was used as a toy example for a different purpose (Nakagawa & Schielzeth 2013). It offers a balanced dataset with rather simple structure, sizable effects and decent sample size, just right for demonstrating some features of rptR. Sufficient sample size is required in particular for the non-Gaussian traits, because those tend to be more computationally demanding and less rich in information per data point than simple Gaussian traits.

In brief the imaginary sampling design of the simulated dataset is as follows. Beetle larvae were sampled from 12 populations (‘Population’) with samples taken from two discrete microhabitats at each location (‘Habitat’). Samples were split in equal proportion and raised in two dietary treatments (‘Treatment’). Beetles were sexed at the pupal stage (‘Sex’) and pupae were kept in sex-homogeneous containers (‘Container’). The phenotype in this dataset is body length (‘BodyL’).

References


BeetlesFemale dataset

Details

This is an simulated dataset which was used as a toy example for a different purpose (Nakagawa & Schielzeth 2013). It offers a balanced dataset with rather simple structure, sizable effects and decent sample size, just right for demonstrating some features of rptR. Sufficient sample size is required in particular for the non-Gaussian traits, because those tend to be more computationally demanding and less rich in information per data point than simple Gaussian traits.
In brief the imaginary sampling design of the simulated dataset is as follows. Beetle larvae were sampled from 12 populations (‘Population’) with samples taken from two discrete microhabitats at each location (‘Habitat’). Samples were split in equal proportion and raised in two dietary treatments (‘Treatment’). Beetles were sexed at the pupal stage (‘Sex’) and pupae were kept in sex-homogeneous containers (‘Container’). The phenotype in this dataset is the number of eggs laid by female beetles (‘Egg’).

References


---

BeetlesMale

**BeetlesMale dataset**

**Description**

BeetlesMale dataset

**Details**

This is an simulated dataset which was used as a toy example for a different purpose (Nakagawa & Schielzeth 2013). It offers a balanced dataset with rather simple structure, sizable effects and decent sample size, just right for demonstrating some features of rptR. Sufficient sample size is required in particular for the non-Gaussian traits, because those tend to be more computationally demanding and less rich in information per data point than simple Gaussian traits.

In brief the imaginary sampling design of the simulated dataset is as follows. Beetle larvae were sampled from 12 populations (‘Population’) with samples taken from two discrete microhabitats at each location (‘Habitat’). Samples were split in equal proportion and raised in two dietary treatments (‘Treatment’). Beetles were sexed at the pupal stage (‘Sex’) and pupae were kept in sex-homogeneous containers (‘Container’). The phenotype in this dataset is a binary variable containing the two distinct color morphs of males: dark and reddish-brown (‘Colour’).

References

plot.rpt

Plot a rpt object

Description

Plots the distribution of repeatability estimates from bootstrapping and permutation tests.

Usage

```r
## S3 method for class 'rpt'
plot(x, grname = names(x$ngroups), scale = c("link", "original"), type = c("boot", "permut"), main = NULL, breaks = "FD", xlab = NULL, ...)
```

Arguments

- `x`: An rpt object returned from one of the rpt functions.
- `grname`: The name of the grouping factor to plot.
- `scale`: Either "link" or "original" scale results for results of non-Gaussian functions.
- `type`: Either "boot" or "permut" for plotting the results of bootstraps or permutations.
- `main`: Plot title
- `breaks`: hist() argument
- `xlab`: x-axis title

Value

A histogram of the distribution of bootstrapping or permutation test estimates of the repeatability including a confidence interval (CI).

Author(s)

Holger Schielzeth (holger.schielzeth@uni-jena.de), Shinichi Nakagawa (s.nakagawa@unsw.edu.au), Martin Stoffel (martin.adam.stoffel@gmail.com)

References

print.rpt

Print a rpt object

Description

Displays the results a rpt object (i.e. the result of a rpt function call) in a nice form.

Usage

```r
## S3 method for class 'rpt'
print(x, ...)
```

Arguments

- `x` An rpt object returned from one of the rpt functions
- `...` Additional arguments; none are used in this method.

Value

Abbreviations in the print.rpt output:

- **R**: Repeatability.
- **SE**: Standard error of R.
- **CI**: Confidence interval of R derived from parametric bootstrapping.
- **P**: P-value
- **LRT**: Likelihood-ratio test
- **Permutation**: Permutation of residuals

Author(s)

Holger Schielzeth (holger.schielzeth@uni-jena.de), Shinichi Nakagawa (s.nakagawa@unsw.edu.au), Martin Stoffel (martin.adam.stoffel@gmail.com)

References

print.summary.rpt  Prints the summary of a rpt object

Description
Displays the summary of an rpt object (i.e. the result of a rpt function call) in an extended form.

Usage
## S3 method for class 'summary.rpt'
print(x, ...)

Arguments
- **x**: An rpt object returned from one of the rpt functions
- **...**: Additional arguments; none are used in this method.

Author(s)
Holger Schielzeth (holger.schielzeth@uni-jena.de), Shinichi Nakagawa (s.nakagawa@unsw.edu.au), Martin Stoffel (martin.adam.stoffel@gmail.com)

References

rpt  Repeatability Estimation for Gaussian and Non-Gaussian Data

Description
A wrapper function for (adjusted) repeatability estimation from generalized linear mixed-effects models fitted by restricted maximum likelihood (REML). Calls specialised functions depending of the choice of datatype and method.

Usage
rpt(formula, grname, data, datatype = c("Gaussian", "Binary", "Proportion", "count"), link = c("logit", "probit", "log", "sqrt"), CI = 0.95, n.boot = 1000, n.permut = 0, parallel = FALSE, n.cores = NULL, ratio = TRUE, adjusted = TRUE, expect = "meanobs", rpt.obj = NULL, update = FALSE)
Arguments

formula Formula as used e.g. by lmer. The grouping factor(s) of interest needs to be included as a random effect, e.g. `(1|groups)`. Covariates and additional random effects can be included to estimate adjusted repeatabilities.

gname A character string or vector of character strings giving the name(s) of the grouping factor(s), for which the repeatability should be estimated. Spelling needs to match the random effect names as given in formula and terms have to be set in quotation marks. The reserved terms "Residual", "Overdispersion" and "Fixed" allow the estimation of overdispersion variance, residual variance and variance explained by fixed effects, respectively.

data A dataframe that contains the variables included in the formula and gname arguments.

datatype Character string specifying the data type ("Gaussian", "Binary", "Proportion", "Poisson").

link Character string specifying the link function. Ignored for "Gaussian" datatype.

CI Width of the required confidence interval between 0 and 1 (defaults to 0.95).

nboot Number of parametric bootstraps for interval estimation (defaults to 1000). Larger numbers of bootstraps give a better asymptotic CI, but may be time-consuming. Bootstrapping can be switch off by setting nboot = 0. See also Details below.

npermut Number of permutations used when calculating asymptotic p-values (defaults to 0). Larger numbers of permutations give a better asymptotic p-values, but may be time-consuming (in particular when multiple grouping factors are specified). Permutaton tests can be switch off by setting npermut = 0. See also Details below.

parallel Boolean to express if parallel computing should be applied (defaults to FALSE). If TRUE, bootstraps and permutations will be distributed across multiple cores.

ncores Specifying the number of cores to use for parallelization. On default, all but one of the available cores are used.

ratio Boolean to express if variances or ratios of variance should be estimated. If FALSE, the variance(s) are returned without forming ratios. If TRUE (the default) ratios of variances (i.e. repeatabilities) are estimated.

adjusted Boolean to express if adjusted or unadjusted repeatabilities should be estimated. If TRUE (the default), the variances explained by fixed effects (if any) will not be part of the denominator, i.e. repeatabilities are calculated after controlling for variation due to covariates. If FALSE, the variances explained by fixed effects (if any) will be added to the denominator.

expect A character string specifying the method for estimating the expectation in Poisson models with log link and in Binomial models with logit link (in all other cases the agrument is ignored). The only valid terms are 'meanobs' and 'latent' (and 'liability for binary and proportion data). With the default 'meanobs', the expectation is estimated as the mean of the observations in the sample. With 'latent', the expectation is estimated from estimates of the intercept and variances on the link scale. While this is a preferred solution, it is susceptible to the distribution of fixed effect covariates and gives appropriate results typically
only when all covariances are centered to zero. With 'liability’ estimates follow
formulae as presented in Nakagawa & Schielzeth (2010). Liability estimates
tend to be slightly higher.

rptObj
The output of a rptR function. Can be specified in combination with update =
TRUE to update bootstraps and permutations

update
If TRUE, the rpt object to be updated has to be inputted with the rptObj argu-
ment. The function just updates the permutations and bootstraps, so make sure
to specify all other arguments exactly like for the rpt object specified in rptObj.

Details
For datatype='Gaussian' calls function rptGaussian, for datatype='Poisson' calls function
rptPoisson, for datatype='Binary' calls function rptBinary, for datatype='Proportion' calls
function rptProportion.

Confidence intervals and standard errors are estimated by parametric bootstrapping. Under
the assumption that the model is specified correctly, the fitted model can be used to generate response
values that could potentially be observed. Differences between the original data and the simulated
response from the fitted model arise from sampling variation. The full model is then fitted to each
simulated response vector. The distribution of estimates across all nboot replicates represents the
design- and model-specific sampling variance and hence uncertainty of the estimates.

In addition to the likelihood-ratio test, the package uses permutation tests for null hypothesis test-
ing. The general idea is to randomize data under the null hypothesis of no effect and then test in
how many cases the estimates from the model reach or exceed those in the observed data. In the
simplest case, a permutation test randomizes the vector of group identities against the response vec-
tor many times, followed by refitting the model and recalculating the repeatabilities. This provides
a null distribution for the case that group identities are unrelated to the response. However, in more
complex models involving multiple random effects and/or fixed effects, such a procedure will also
break the data structure between the grouping factor of interest and other aspects of the experimental
design. Therefore rptR implements a more robust alternative which works by fitting a model with-
out the grouping factor of interest. It then adds the randomized residuals to the fitted values of this
model, followed by recalculating the repeatability from the full model. This procedure maintains
the general data structure and any effects other than the grouping effect of interest. The number of
permutations can be adjusted with the nperm argument. By the logic of a null hypothesis testing, the
observed data is one possible (albeit maybe unlikely) outcome under the null hypothesis. So the
observed data is always included as one ‘randomization’ and the P value can thus never be lower
than 1/nperm, because at least one randomization is as extreme as the observed data.

Note also that the likelihood-ratio test, since testing variances at the boundary of the possible
parameter range (i.e. against zero), uses a mixture distribution of Chi-square distributions with zero
and one degree of freedom as a reference. This is equivalent to deviding the P value derived from
a Chi-square distribution with one degree of freedom by two.

Value
Returns an object of class rpt. See specific functions for details.
Author(s)
Holger Schielzeth (holger.schielzeth@uni-jena.de), Shinichi Nakagawa (s.nakagawa@unsw.edu.au),
Martin Stoffel (martin.adam.stoffel@gmail.com)

References
Nakagawa, S. & Schielzeth, H. (2010) Repeatability for Gaussian and non-Gaussian data: a prac-

See Also
rptR

Examples
# load data
data(BeetlesBody)
data(BeetlesMale)
data(BeetlesFemale)

# prepare proportion data
BeetlesMale$Dark <- BeetlesMale$Colour
BeetlesMale$Reddish <- (BeetlesMale$Colour-1)*-1
BeetlesColour <- aggregate(cbind(Dark, Reddish) ~ Treatment + Population + Container,
data=BeetlesMale, FUN=sum)

# Note: nboot and npermut are set to 0 for speed reasons. Use larger numbers
# for the real analysis.

# gaussian data (example with a single random effect)
rpt(BodyL ~ (1|Population), grname="Population", data=BeetlesBody,
nboot=0, npermut=0, datatype = "Gaussian")

# poisson data (example with two grouping levels and adjusted for fixed effect)
rpt(Egg ~ Treatment + (1|Container) + (1|Population), grname=c("Population"),
data = BeetlesFemale, nboot=0, npermut=0, datatype = "Poisson")

## Not run:

# binary data (example with estimation of the fixed effect variance)
rpt(Colour ~ Treatment + (1|Container) + (1|Population),
grname=c("Population", "Container", "Fixed"),
data=BeetlesMale, nboot=0, npermut=0, datatype = "Binary", adjusted = FALSE)

# proportion data (example for the estimation of raw variances,
# including residual and fixed-effect variance)
rpt(cbind(Dark, Reddish) ~ Treatment + (1|Population),
grname=c("Population", "Residual", "Fixed"), data=BeetlesColour,
nboot=0, npermut=0, datatype = "Proportion", ratio=FALSE)

## End(Not run)
rptBinary

GLMM-based Repeatability Estimation for Binary Data

Description

Estimates repeatability from a generalized linear mixed-effects models fitted by restricted maximum likelihood (REML).

Usage

rptBinary(formula, grname, data, link = c("logit", "probit"), CI = 0.95, 
nboot = 1000, npermut = 0, parallel = FALSE, ncores = NULL, 
ratio = TRUE, adjusted = TRUE, expect = "meanobs", rptObj = NULL, 
update = FALSE)

Arguments

formula  Formula as used e.g. by lmer. The grouping factor(s) of interest needs to be included as a random effect, e.g. '(1|groups)'. Covariates and additional random effects can be included to estimate adjusted repeatabilities.

grname   A character string or vector of character strings giving the name(s) of the grouping factor(s), for which the repeatability should be estimated. Spelling needs to match the random effect names as given in formula and terms have to be set in quotation marks. The reserved terms "Residual", "Fixed" allow the estimation of residual variance and variance explained by fixed effects, respectively. "Overdispersion" is not available for rptBinary.

data     A dataframe that contains the variables included in the formula and grname arguments.

link     Link function. logit and probit are allowed, defaults to logit.

CI       Width of the required confidence interval between 0 and 1 (defaults to 0.95).

nboot    Number of parametric bootstraps for interval estimation (defaults to 1000). Larger numbers of bootstraps give a better asymptotic CI, but may be time-consuming. Bootstrapping can be switch off by setting nboot = 0. See also Details below.

npermut  Number of permutations used when calculating asymptotic p-values (defaults to 0). Larger numbers of permutations give a better asymptotic p-values, but may be time-consuming (in particular when multiple grouping factors are specified). Permutation tests can be switch off by setting npermut = 0. See also Details below.

parallel Boolean to express if parallel computing should be applied (defaults to FALSE). If TRUE, bootstraps and permutations will be distributed across multiple cores.

ncores   Specifying the number of cores to use for parallelization. On default, all but one of the available cores are used.
**rptBinary**

- **ratio**
  - Boolean to express if variances or ratios of variance should be estimated. If FALSE, the variance(s) are returned without forming ratios. If TRUE (the default) ratios of variances (i.e. repeatabilities) are estimated.

- **adjusted**
  - Boolean to express if adjusted or unadjusted repeatabilities should be estimated. If TRUE (the default), the variances explained by fixed effects (if any) will not be part of the denominator, i.e. repeatabilities are calculated after controlling for variation due to covariates. If FALSE, the variances explained by fixed effects (if any) will be added to the denominator.

- **expect**
  - A character string specifying the method for estimating the expectation in Poisson models with log link and in Binomial models with logit link (in all other cases the argument is ignored). The only valid terms are 'meanobs' and 'latent' (and 'liability for binary and proportion data). With the default 'meanobs', the expectation is estimated as the mean of the observations in the sample. With 'latent', the expectation is estimated from estimates of the intercept and variances on the link scale. While this is a preferred solution, it is susceptible to the distribution of fixed effect covariates and gives appropriate results typically only when all covariates are centered to zero. With 'liability' estimates follow formulae as presented in Nakagawa & Schielzeth (2010). Liability estimates tend to be slightly higher.

- **rptObj**
  - The output of a rptR function. Can be specified in combination with update = TRUE to update bootstraps and permutations.

- **update**
  - If TRUE, the rpt object to be updated has to be inputted with the rptObj argument. The function just updates the permutations and bootstraps, so make sure to specify all other arguments exactly like for the rpt object specified in rptObj.

**Details**

see details section of rpt for details on parametric bootstrapping, permutation and likelihood-ratio tests.

**Value**

Returns an object of class rpt that is a list with the following elements:

- **call**
  - Function call.

- **datatype**
  - Response distribution (here: 'Binary').

- **CI**
  - Coverage of the confidence interval as specified by the CI argument.

- **R**
  - data.frame with point estimates for repeatabilities. Columns represent grouping factors of interest. Rows show original and link scale repeatabilities (in this order).

- **se**
  - data.frame with approximate standard errors (se) for repeatabilities. Columns are groups of interest. Rows are original and link scale (in this order). Note that the distribution might not be symmetrical, in which case the emphasis is less informative.

- **CI_emp**
  - list of two elements containing the confidence intervals for repeatabilities on the link and original scale, respectively. Within each list element, lower and upper CI are columns and each row for each grouping factor of interest.
data frame with p-values from a significance test based on likelihood-ratios in the first column and significance test based on permutation of residuals for both the original and link scale in the second and third column. Each row represents a grouping factor of interest.

Parametric bootstrap samples for R on the link scale. Each list element is a grouping factor.

Parametric bootstrap samples for R on the original scale. Each list element is a grouping factor.

Permutation samples for R on the link scale. Each list element is a grouping factor.

Permutation samples for R on the original scale. Each list element is a grouping factor.

List of two elements. LRT_mod is the likelihood for the full model and (2) LRT_table is a data.frame for the reduced model(s) including columns for the likelihood logl_red, the likelihood ratio(s) LR_D, p-value(s) LR_P and degrees of freedom for the likelihood-ratio test(s) LR_df.

Number of groups for each grouping level.

Number of observations.

Fitted model.

Boolean. TRUE, if ratios have been estimated, FALSE, if variances have been estimated

Boolean. TRUE, if estimates are adjusted

list with two elements. 'warnings_boot' and 'warnings_permut' contain warnings from the lme4 model fitting of bootstrap and permutation samples, respectively.

Holger Schielzeth (holger.schielzeth@uni-jena.de), Shinichi Nakagawa (s.nakagawa@unsw.edu.au) & Martin Stoffel (martin.adam.stoffel@gmail.com)


See Also

rpt
Examples

data(BeetlesMale)

# Note: nboot and npermut are set to 0 for speed reasons.

# repeatability with one grouping level
rptBinary(Colour ~ (1|Population), grname=c("Population"),
data=BeetlesMale, nboot=0, npermut=0)

# unadjusted repeatabilities with fixed effects and
# estimation of the fixed effect variance
rptBinary(Colour ~ Treatment + (1|Container) + (1|Population),
grname=c("Container", "Population", "Fixed"),
data=BeetlesMale, nboot=0, npermut=0, adjusted=FALSE)

## Not run:
# variance estimation of random effects and residual
R_est <- rptBinary(Colour ~ Treatment + (1|Container) + (1|Population),
grname=c("Container", "Population", "Residual"),
data = BeetlesMale, nboot=0, npermut=0, ratio = FALSE)

## End(Not run)

---

**rptGaussian**  
*LMM-based Repeatability Estimation for Gaussian Data*

**Description**

Estimates the repeatability from a general linear mixed-effects models fitted by restricted maximum likelihood (REML).

**Usage**

```r
rptGaussian(formula, grname, data, CI = 0.95, nboot = 1000, npermut = 0,
  parallel = FALSE, ncores = NULL, ratio = TRUE, adjusted = TRUE,
  rptObj = NULL, update = FALSE)
```

**Arguments**

- **formula**: Formula as used e.g. by *lmer*. The grouping factor(s) of interest needs to be included as a random effect, e.g. `(1|groups)`. Covariates and additional random effects can be included to estimate adjusted repeatabilities.

- **grname**: A character string or vector of character strings giving the name(s) of the grouping factor(s), for which the repeatability should be estimated. Spelling needs to match the random effect names as given in `formula` and terms have to be set in
quotation marks. The reserved terms "Residual", "Overdispersion" and "Fixed" allow the estimation of overdispersion variance, residual variance and variance explained by fixed effects, respectively.

data A dataframe that contains the variables included in the formula and grname arguments.

CI Width of the required confidence interval between 0 and 1 (defaults to 0.95).

nboot Number of parametric bootstraps for interval estimation (defaults to 1000). Larger numbers of bootstraps give a better asymptotic CI, but may be time-consuming. Bootstrapping can be switch off by setting nboot = 0. See also Details below.

npermut Number of permutations used when calculating asymptotic p-values (defaults to 0). Larger numbers of permutations give a better asymptotic p-values, but may be time-consuming (in particular when multiple grouping factors are specified). Permutation tests can be switch off by setting npermut = 0. See also Details below.

parallel Boolean to express if parallel computing should be applied (defaults to FALSE). If TRUE, bootstraps and permutations will be distributed across multiple cores.

ncores Specifying the number of cores to use for parallelization. On default, all but one of the available cores are used.

ratio Boolean to express if variances or ratios of variance should be estimated. If FALSE, the variance(s) are returned without forming ratios. If TRUE (the default) ratios of variances (i.e. repeatabilities) are estimated.

adjusted Boolean to express if adjusted or unadjusted repeatabilities should be estimated. If TRUE (the default), the variances explained by fixed effects (if any) will not be part of the denominator, i.e. repeatabilities are calculated after controlling for variation due to covariates. If FALSE, the variances explained by fixed effects (if any) will be added to the denominator.

rptObj The output of a rptR function. Can be specified in combination with update = TRUE to update bootstraps and permutations.

update If TRUE, the rpt object to be updated has to be inputted with the rptObj argument. The function just updates the permutations and bootstraps, so make sure to specify all other arguments exactly like for the rpt object specified in rptObj.

Details

see details section of rpt for details on parametric bootstrapping, permutation and likelihood-ratio tests.

Value

Returns an object of class rpt that is a a list with the following elements:

call Function call.
datatype Response distribution (here: 'Gaussian').
CI Coverage of the confidence interval as specified by the CI argument.
R data.frame with point estimates for repeatabilities for each grouping factor of interest.
se

data.frame with approximate standard errors (se) for repeatabilities. Rows represent grouping factors of interest. Note that the distribution might not be symmetrical, in which case the se is less informative.

CI_emp

data.frame containing the (empirical) confidence intervals for the repeatabilities estimated based parametric bootstrapping. Each row represents a grouping factor of interest.

P

data.frame with p-values based on likelihood-ratio tests (first column) and permutation tests (second column). Each row represents a grouping factor of interest.

R_boot

Vector(s) of parametric bootstrap samples for R. Each list element represents a grouping factor.

R_permut

Vector(s) of permutation samples for R. Each list element represents a grouping factor.

LRT

list with two elements. (1) The likelihood for the full model and a data.frame called LRT_table for the reduced model(s), which includes columns for the respective grouping factor(s), the likelihood(s) logL_red, likelihood ratio(s) LR_D, p-value(s) LRT_P and degrees of freedom LRT_df

groups

Number of groups for each grouping level.

nobs

Number of observations.

mod

Fitted model.

ratio

Boolean. TRUE, if ratios have been estimated, FALSE, if variances have been estimated

adjusted

Boolean. TRUE, if estimates are adjusted

all_warnings

list with two elements. 'warnings_boot' and 'warnings_permut' contain warnings from the lme4 model fitting of bootstrap and permutation samples, respectively.

Author(s)

Holger Schielzeth (holger.schielzeth@uni-jena.de), Shinichi Nakagawa (s.nakagawa@unsw.edu.au) & Martin Stoffel (martin.adam.stoffel@gmail.com)

References


See Also

rpt
rptPoisson

GLMM-based Repeatability Estimation for Poisson-distributed Data

Description

Estimates repeatability from a generalized linear mixed-effects models fitted by restricted maximum likelihood (REML).

Usage

rptPoisson(formula, grname, data, link = c("log", "sqrt"), CI = 0.95,
  nboot = 1000, npermut = 0, parallel = FALSE, ncores = NULL,
  ratio = TRUE, adjusted = TRUE, expect = "meanobs", rptObj = NULL,
  update = FALSE)

Arguments

formula Formua as used e.g. by lmer. The grouping factor(s) of interest needs to be included as a random effect, e.g. '(1|groups)'. Covariates and additional random effects can be included to estimate adjusted repeatabilities.
grname
A character string or vector of character strings giving the name(s) of the grouping factor(s), for which the repeatability should be estimated. Spelling needs to match the random effect names as given in formula and terms have to be set in quotation marks. The reserved terms "Residual", "Overdispersion" and "Fixed" allow the estimation of overdispersion variance, residual variance and variance explained by fixed effects, respectively.

data
A dataframe that contains the variables included in the formula and grname arguments.

link
Link function. logit and probit are allowed, defaults to logit.

CI
Width of the required confidence interval between 0 and 1 (defaults to 0.95).

nboot
Number of parametric bootstraps for interval estimation (defaults to 1000). Larger numbers of bootstraps give a better asymptotic CI, but may be time-consuming. Bootstrapping can be switch off by setting nboot = 0. See also Details below.

npermute
Number of permutations used when calculating asymptotic p-values (defaults to 0). Larger numbers of permutations give a better asymptotic p-values, but may be time-consuming (in particular when multiple grouping factors are specified). Permutation tests can be switch off by setting npermute = 0. See also Details below.

parallel
Boolean to express if parallel computing should be applied (defaults to FALSE). If TRUE, bootstraps and permutations will be distributed across multiple cores.

ncore
Specifying the number of cores to use for parallelization. On default, all but one of the available cores are used.

ratio
Boolean to express if variances or ratios of variance should be estimated. If FALSE, the variance(s) are returned without forming ratios. If TRUE (the default) ratios of variances (i.e. repeatabilities) are estimated.

adjusted
Boolean to express if adjusted or unadjusted repeatabilities should be estimated. If TRUE (the default), the variances explained by fixed effects (if any) will not be part of the denominator, i.e. repeatabilities are calculated after controlling for variation due to covariates. If FALSE, the variances explained by fixed effects (if any) will be added to the denominator.

expect
A character string specifying the method for estimating the expectation in Poisson models with log link and in Binomial models with logit link (in all other cases the argument is ignored). The only valid terms are 'meanobs' and 'latent' (and 'liability for binary and proportion data). With the default 'meanobs', the expectation is estimated as the mean of the observations in the sample. With 'latent', the expectation is estimated from estimates of the intercept and variances on the link scale. While this is a preferred solution, it is susceptible to the distribution of fixed effect covariates and gives appropriate results typically only when all covariances are centered to zero. With 'liability' estimates follow formulae as presented in Nakagawa & Schielzeth (2010). Liability estimates tend to be slightly higher.

rptObj
The output of a rptR function. Can be specified in combination with update = TRUE to update bootstraps and permutations.

update
If TRUE, the rpt object to be updated has to be inputted with the rptObj argument. The function just updates the permutations and bootstraps, so make sure to specify all other arguments exactly like for the rpt object specified in rptObj.
Details

see details section of `rpt` for details on parametric bootstrapping, permutation and likelihood-ratio tests.

Value

Returns an object of class `rpt` that is a a list with the following elements:

- **call**: Function call
- **datatype**: Response distribution (here: 'Poisson').
- **CI**: Coverage of the confidence interval as specified by the CI argument.
- **R**: `data.frame` with point estimates for repeatabilities. Columns represent grouping factors of interest. Rows show original and link scale repeatabilities (in this order).
- **se**: `data.frame` with approximate standard errors (`se`) for repeatabilities. Columns are groups of interest. Rows are original and link scale (in this order). Note that the distribution might not be symmetrical, in which case the `se` is less informative.
- **CI_emp**: list of two elements containing the confidence intervals for repeatabilities on the link and original scale, respectively. Within each list element, lower and upper CI are columns and each row for each grouping factor of interest.
- **P**: `data.frame` with p-values from a significance test based on likelihood-ratios in the first column and significance test based on permutation of residuals for both the original and link scale in the second and third column. Each row represents a grouping factor of interest.
- **R_boot_link**: Parametric bootstrap samples for `R` on the link scale. Each list element is a grouping factor.
- **R_boot_org**: Parametric bootstrap samples for `R` on the original scale. Each list element is a grouping factor.
- **R_permut_link**: Permutation samples for `R` on the link scale. Each list element is a grouping factor.
- **R_permut_org**: Permutation samples for `R` on the original scale. Each list element is a grouping factor.
- **LRT**: List of two elements. `LRT_mod` is the likelihood for the full model and (2) `LRT_table` is a data.frame for the reduced model(s) including columns for the likelihood `logl_red`, the likelihood ratio(s) `LR_D`, p-value(s) `LR_P` and degrees of freedom for the likelihood-ratio test(s) `LR_df`.
- **ngroups**: Number of groups for each grouping level.
- **nobs**: Number of observations.
- **mod**: Fitted model.
- **ratio**: Boolean. TRUE, if ratios have been estimated, FALSE, if variances have been estimated
- **adjusted**: Boolean. TRUE, if estimates are adjusted
- **all_warnings**: list with two elements. 'warnings_boot' and 'warnings_permut' contain warnings from the lme4 model fitting of bootstrap and permutation samples, respectively.
Author(s)

Holger Schielzeth (holger.schielzeth@uni-jena.de), Shinichi Nakagawa (s.nakagawa@unsw.edu.au) & Martin Stoffel (martin.adam.stoffel@gmail.com)

References


See Also

rpt

Examples

```r
# load data
data(BeetlesFemale)

# Note: nboot and npermut are set to 0 for speed reasons.

# estimating adjusted repeatabilities for two random effects
rptPoisson(Egg ~ Treatment + (1|Container) + (1|Population),
            grname=c("Container", "Population"),
            data = BeetlesFemale, nboot=0, npermut=0)

# unadjusted repeatabilities with fixed effects and
# estimation of the fixed effect variance
rptPoisson(Egg ~ Treatment + (1|Container) + (1|Population),
            grname=c("Container", "Population", "Fixed"),
            data=BeetlesFemale, nboot=0, npermut=0, adjusted=FALSE)

# variance estimation of random effects, residual and overdispersion
rptPoisson(formula = Egg ~ Treatment + (1|Container) + (1|Population),
            grname=c("Container","Population","Residual", "Overdispersion"),
            data = BeetlesFemale, nboot=0, npermut=0, ratio = FALSE)
```
rptProportion

GLMM-based Repeatability Estimation for Proportion Data

Description

Estimates repeatability from a generalized linear mixed-effects models fitted by restricted maximum likelihood (REML).

Usage

rptProportion(formula, grname, data, link = c("logit", "probit"), CI = 0.95, 
nboot = 1000, nperm = 0, parallel = FALSE, ncores = NULL, 
ratio = TRUE, adjusted = TRUE, expect = "meanobs", rptobj = NULL, 
update = FALSE)

Arguments

formula Formula as used e.g. by lmer. The grouping factor(s) of interest needs to be included as a random effect, e.g. '(1|groups)'. Covariates and additional random effects can be included to estimate adjusted repeatabilities.

grname A character string or vector of character strings giving the name(s) of the grouping factor(s), for which the repeatability should be estimated. Spelling needs to match the random effect names as given in formula and terms have to be set in quotation marks. The reserved terms "Residual", "Overdispersion" and "Fixed" allow the estimation of overdispersion variance, residual variance and variance explained by fixed effects, respectively.

data A dataframe that contains the variables included in the formula and grname arguments.

link Link function. logit and probit are allowed, defaults to logit.

CI Width of the required confidence interval between 0 and 1 (defaults to 0.95).

nboot Number of parametric bootstraps for interval estimation (defaults to 1000). Larger numbers of bootstraps give a better asymptotic CI, but may be time-consuming. Bootstrapping can be switch off by setting nboot = 0. See also Details below.

nperm Number of permutations used when calculating asymptotic p-values (defaults to 0). Larger numbers of permutations give a better asymptotic p-values, but may be time-consuming (in particular when multiple grouping factors are specified). Permuatation tests can be switch off by setting nperm = 0. See also Details below.

parallel Boolean to express if parallel computing should be applied (defaults to FALSE). If TRUE, bootstraps and permutations will be distributed across multiple cores.

ncores Specifying the number of cores to use for parallelization. On default, all but one of the available cores are used.

ratio Boolean to express if variances or ratios of variance should be estimated. If FALSE, the variance(s) are returned without forming ratios. If TRUE (the default) ratios of variances (i.e. repeatabilities) are estimated.
adjusted

Boolean to express if adjusted or unadjusted repeatabilities should be estimated. If TRUE (the default), the variances explained by fixed effects (if any) will not be part of the denominator, i.e. repeatabilities are calculated after controlling for variation due to covariates. If FALSE, the variances explained by fixed effects (if any) will be added to the denominator.

expect

A character string specifying the method for estimating the expectation in Poisson models with log link and in Binomial models with logit link (in all other cases the argument is ignored). The only valid terms are 'meanobs' and 'latent' (and 'liability for binary and proportion data). With the default 'meanobs', the expectation is estimated as the mean of the observations in the sample. With 'latent', the expectation is estimated from estimates of the intercept and variances on the link scale. While this is a preferred solution, it is susceptible to the distribution of fixed effect covariates and gives appropriate results typically only when all covariances are centered to zero. With 'liability' estimates follow formulae as presented in Nakagawa & Schielzeth (2010). Liability estimates tend to be slightly higher.

rptObj

The output of a rptR function. Can be specified in combination with update = TRUE to update bootstraps and permutations

update

If TRUE, the rpt object to be updated has to be inputted with the rptObj argument. The function just updates the permutations and bootstraps, so make sure to specify all other arguments exactly like for the rpt object specified in rptObj.

Details

see details section of rpt for details on parametric bootstrapping, permutation and likelihood-ratio tests.

Value

Returns an object of class rpt that is a a list with the following elements:

call

Function call

datatype

Response distribution (here: 'Proportion').

CI

Width of the confidence interval.

R

data.frame with point estimates for repeatabilities. Columns represent grouping factors of interest. Rows show original and link scale repeatabilities (in this order).

se

data.frame with approximate standard errors (se) for repeatabilities. Columns are groups of interest. Rows are original and link scale (in this order). Note that the distribution might not be symmetrical, in which case the se is less informative.

CI_emp

list of two elements containing the confidence intervals for repeatabilities on the link and original scale, respectively. Within each list element, lower and upper CI are columns and each row for each grouping factor of interest.

P

data.frame with p-values from a significance test based on likelihood-ratios in the first column and significance test based on permutation of residuals for both the original and link scale in the second and third column. Each row represents a grouping factor of interest.
rptProportion

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>R_boot_link</td>
<td>Parametric bootstrap samples for $R$ on the link scale. Each list element is a grouping factor.</td>
</tr>
<tr>
<td>R_boot_org</td>
<td>Parametric bootstrap samples for $R$ on the original scale. Each list element is a grouping factor.</td>
</tr>
<tr>
<td>R_permut_link</td>
<td>Permutation samples for $R$ on the link scale. Each list element is a grouping factor.</td>
</tr>
<tr>
<td>R_permut_org</td>
<td>Permutation samples for $R$ on the original scale. Each list element is a grouping factor.</td>
</tr>
<tr>
<td>LRT</td>
<td>List of two elements. $LRT_mod$ is the likelihood for the full model and (2) $LRT_table$ is a data.frame for the reduced model(s) including columns for the likelihood $\log l_red$, the likelihood ratio(s) $LR_D$, p-value(s) $LR_P$ and degrees of freedom for the likelihood-ratio test(s) $LR_df$.</td>
</tr>
<tr>
<td>ngroups</td>
<td>Number of groups for each grouping level.</td>
</tr>
<tr>
<td>nobs</td>
<td>Number of observations.</td>
</tr>
<tr>
<td>overdisp</td>
<td>Overdispersion parameter. Equals the variance in the observational factor random effect</td>
</tr>
<tr>
<td>mod</td>
<td>Fitted model.</td>
</tr>
<tr>
<td>ratio</td>
<td>Boolean. TRUE, if ratios have been estimated, FALSE, if variances have been estimated</td>
</tr>
<tr>
<td>adjusted</td>
<td>Boolean. TRUE, if estimates are adjusted</td>
</tr>
<tr>
<td>all_warnings</td>
<td>list with two elements. 'warnings_boot' and 'warnings_permut' contain warnings from the lme4 model fitting of bootstrap and permutation samples, respectively.</td>
</tr>
</tbody>
</table>

Author(s)

Holger Schielzeth (holger.schielzeth@uni-jena.de), Shinichi Nakagawa (s.nakagawa@unsw.edu.au) & Martin Stoffel (martin.adam.stoffel@gmail.com)

References


See Also

rpt
Examples

data(BeetlesMale)

# prepare proportion data
BeetlesMale$Dark <- BeetlesMale$Colour
BeetlesMale$Reddish <- (BeetlesMale$Colour - 1) * -1
BeetlesColour <- aggregate(cbind(Dark, Reddish) ~ Treatment + Population + Container, 
data=BeetlesMale, FUN=sum)

# Note: nboot and npermut are set to 0 for speed reasons.

# repeatability with one grouping level
rptProportion(cbind(Dark, Reddish) ~ (1|Population), 
grname=c("Population"), data=BeetlesColour, nboot=3, npermut=3)

# unadjusted repeatabilities with fixed effects and
# estimation of the fixed effect variance
rptProportion(cbind(Dark, Reddish) ~ Treatment + (1|Container) + (1|Population), 
grname=c("Population", "Fixed"), 
data=BeetlesColour, nboot=0, npermut=0, adjusted=FALSE)

# variance estimation of random effects, residual and overdispersion
rptProportion(cbind(Dark, Reddish) ~ Treatment + (1|Container) + (1|Population), 
grname=c("Container", "Reddish", "Population", "Residual", "Overdispersion"), 
data = BeetlesColour, nboot=0, npermut=0, ratio = FALSE)

---

**rptR**

**rptR: Repeatability Estimation for Gaussian and Non-Gaussian data**

Description

A collection of functions for calculating point estimates, interval estimates and significance tests of the repeatability (intra-class correlation coefficient) as well as variance components in mixed effects models. The function **rpt** is a wrapper function that calls more specialised functions as required. Specialised functions can also be called directly (see **rpt** for details). All functions return lists of values in the form of an S3 object **rpt**. The function **summary.rpt** produces summaries in a detailed format and **plot.rpt** plots bootstraps or permutation results.

Note

Currently there four different functions depending on the distribution and type of response: (1) **rptGaussian** for a Gaussian response distributions, (2) **rptPoisson** for Poisson-distributed data, (3) **rptBinary** for binary response following binomial distributions and (4) **rptProportion** for response matrices with a column for successes and a column for failures that are analysed as proportions following binomial distributions. All function use a mixed model framework in **lme4**, and the non-Gaussian functions use an observational level random effect to account for overdispersion.

All functions use the argument **formula**, which is the same formula interface as in the **lme4** package (indeed models are fitted by **lmer** or **glmer**). Repeatabilities are calculated for the response
variable, while one or more grouping factors of interest can be assigned as random effects in the form (1|group) and have to be specified with the grname argument. This allows to estimate adjusted repeatabilities (controlling for fixed effects) and the estimation of multiple variance components simultaneously (multiple random effects). All variables have to be columns in a data frame given in the data argument. The link argument specifies the link function for a given non-Gaussian distribution.

The argument ratio allows switching to raw variances rather than ratios of variances to be estimated and The argument adjusted allows switching to an estimation where the variance explained by fixed effects is included in the denominator of the repeatability calculation. The reserved grname terms "Residual", "Overdispersion" and "Fixed" allow the estimation of oversipersion variance, residual variance and variance explained by fixed effects, respectively. All computation can be parallelized with the parallel argument, which enhances computation speed for larger computations.

When using rptR please cite:


Author(s)

Martin Stoffel (martin.adam.stoffel@gmail.com), Shinichi Nakagawa (s.nakagawa@unsw.edu.au) & Holger Schielzeth (holger.schielzeth@uni-jena.de)

References


summary.rpt

Summary of a rpt object

Description

Summary of a rpt object

Usage

## S3 method for class 'rpt'
summary(object, ...)

Arguments

object An rpt object returned from one of the rpt functions
... Additional arguments; none are used in this method.
Author(s)

Holger Schielzeth (holger.schielzeth@uni-jena.de), Shinichi Nakagawa (s.nakagawa@unsw.edu.au),
Martin Stoffel (martin.adam.stoffel@gmail.com)

References

Index

*Topic **datasets**
  BeetlesBody, 2
  BeetlesFemale, 2
  BeetlesMale, 3

*Topic **models**
  plot.rpt, 4
  print.rpt, 5
  print.summary.rpt, 6
  rpt, 6
  summary.rpt, 24

BeetlesBody, 2
BeetlesFemale, 2
BeetlesMale, 3

lmer, 7, 10, 13, 16, 20

plot.rpt, 4, 23
print.rpt, 5
print.summary.rpt, 6

rpt, 6, 11, 12, 14, 15, 18, 19, 21–23
rptBinary, 8, 10, 23
rptGaussian, 8, 13, 23
rptPoisson, 8, 16, 23
rptProportion, 8, 20, 23
rptR, 9, 23
rptR-package (rptR), 23

summary.rpt, 23, 24