Package 'GLMMRR'

January 13, 2021

Type Package

Title Generalized Linear Mixed Model (GLMM) for Binary Randomized
Response Data

Version 0.5.0

Author Jean-Paul Fox [aut], Konrad Klotzke [aut], Duco Veen [aut]

Maintainer Konrad Klotzke <omd.bms.utwente.stats@gmail.com>

Depends R (>= 3.5.0), lme4, methods

Imports lattice, stats, utils, grDevices, RColorBrewer

Description Generalized Linear Mixed Model (GLMM) for Binary Randomized Response Data.
Includes Cauchit, Compl. Log-
Log, Logistic, and Probit link functions for Bernoulli Distributed RR data.
RR Designs: Warner, Forced Response, Unrelated Question, Kuk, Crosswise, and Triangular.
Reference: Fox, J-P, Veen, D. and Klotzke, K. (2018). Generalized Linear Mixed Mod-
els for Randomized Responses. Methodology. <doi:10.1027/1614-2241/a000153>.

License GPL-3

LazyData TRUE

RoxygenNote 7.1.1

NeedsCompilation no

Repository CRAN

Date/Publication 2021-01-13 14:50:03 UTC

R topics documented:

ETHBE   *Online Survey on "Exams and Written Papers"*

Description

The goal of the survey was to estimate the prevalence of various forms of student misconduct such as plagiarizing or cheating in exams. Because students might be reluctant to reveal information on such behaviors, special techniques for sensitive questions were employed in addition to direct questioning. Respondents were randomly assigned to direct questioning or one of five different sensitive question techniques. The dataset contains the (randomized or direct) responses from 4281 students of the University of Bern and ETH Zurich. Each row holds the response to one question for one respondent. The variables are as follows:

Usage

    data(ETHBE)

Format

A data frame in long format with 21405 rows and 29 variables

Details

- id. Identification code of the respondent
- RR_response. Binary randomized or direct response
- Question. Which question was asked
- expcond. Experimental condition

- protect. Level of respondent protection
- subgroup. Subgroups for balanced assignment to experimental conditions
- sample. Sample group
- survey duration. Total time to complete survey (in seconds)
- mobile. Respondent used mobile device (at start of interview)
- java. Javascript version (at start of interview)
- age_cat. Year of birth category
- gender. Gender
- misconduct. Sum score of five binary items on student misconduct
- misconduct2. String of responses to five binary items on student misconduct
- field. Major field of study
- education. Type of study program
- semester. Current semester
- working. Working next to studying
- germanlang. German language skills
- riskattitude. Risk attitude (GSOEP 11-point scale)
- gpa. Current grade point average
- pressure. Studying is a lot of pressure
- stressed. Feeling very stressed in exams
- exams. Number of exams taken
- numberpapers. Number of papers handed in
- RRmodel. Randomized Response Model
- p1. Randomized Response parameter p1
- p2. Randomized Response parameter p2

Author(s)

Marc Hoeglinger, Ben Jann and Andreas Diekmann

References

https://ideas.repec.org/p/bss/wpaper/8.html

---

getCellMeans *Get cell means for unique groups of covariates*

Description

Get cell means for unique groups of covariates

Usage

```
getCellMeans(x, y, factor.groups)
```

Arguments

x	a matrix-like object containing the covariates.
y	a vector of values to compute the means from.
factor.groups	a factor of unique groups of covariates.

Value

the cell means.

---

getCellSizes *Get number of units in each cell*

Description

Get number of units in each cell

Usage

```
getCellSizes(x, n, factor.groups)
```

Arguments

x	a matrix-like object containing the covariates.
n	the total number of units.
factor.groups	a factor of unique groups of covariates.

Value

the number of units in each cell.

---

getMLPrevalence | *Compute Estimated Population Prevalence*

Description

Compute Estimated Population Prevalence

Usage

```
getMLPrevalence(mu, n, c, d)
```

Arguments

mu	observed mean response.
n	number of units.
c	randomized response parameter c.
d	randomized response parameter d.

Value

maximum likelihood estimate of the population prevalence and its variance.

---

getRRparameters | *Compute Randomized Response parameters*

Description

Compute Randomized Response parameters

Usage

```
getRRparameters(vec.RRmodel, vec.p1, vec.p2)
```

Arguments

vec.RRmodel	a character vector of Randomized Response models.
vec.p1	a numeric vector of p1 values.
vec.p2	a numeric vector of p2 values.

Value

a list with c and d values.

---

getUniqueGroups   *Get unique groups of covariates*

Description

Get unique groups of covariates

Usage

```
getUniqueGroups(x)
```

Arguments

x   a matrix-like object containing the covariates.

Value

a factor of unique groups.

MTURK   *MTurk Survey on "Mood and Personality"*

Description

Data from an online validation experiment in which respondents' self-reports of norm breaking behavior were validated against observed actual behavior. After playing a dice game, respondents were asked whether they played honestly, using one of several randomly assigned sensitive question techniques. Furthermore, three other sensitive questions on shoplifting, tax evasion, and voting were asked. The dataset contains the randomized responses from 6152 Amazon Mechanical Turk (MTURK) workers. Each row holds the response to one question for one respondent. The variables are as follows:

Usage

```
data(MTURK)
```

Format

A data frame in long format with 24594 rows and 26 variables

Details

- id. Identification code of the respondent
- Question. Which question was asked
- RR_response. Binary randomized response
- RRp1. Randomized Response parameter p1
- RRp2. Randomized Response parameter p2
- RRmodel. Randomized Response Model
- dicegame. Dice game assignment (1: prediction, 2: roll-a-six)
- cheater. The respondent is classified as honest or cheater if dice game assignment was 'roll-a-six'
- agecategory. Age category
- education. Level of education
- employment. Employment status
- locationinterview. Interview location
- extraversion. Extraversion score on a scale of 2-10
- agreeableness. Agreeableness score on a scale of 2-10
- conscientiousness. Conscientiousness score on a scale of 2-10
- neuroticism. Neuroticism score on a scale of 2-10
- openness. Openness score on a scale of 2-10
- gender. Gender (0: female, 1: male)
- age. Age in years
- privacyquestion1. How well are respondents' anonymity and privacy protected? (1: very poorly, 2: rather poorly, 3: moderately, 4: rather well, 5: very well)
- privacyquestion2. How likely could respondents' sensitive behavior be disclosed by this survey? (1: impossible, 2: not likely, 3: somewhat likely, 4: quite likely, 5: very likely)
- privacyquestion3. Does the special technique absolutely protect your answers? (1: not at all, 2: a little, 3: moderately, 4: quite a bit, 5: definitely)
- privacyquestion4. Do you think you properly followed the instructions for the special technique? (1: not at all, 2: a little, 3: moderately, 4: quite a bit, 5: definitely)
- privacyquestion5. Did you understand how the technique protects respondents? (1: not at all, 2: a little, 3: moderately, 4: quite a bit, 5: definitely)
- region. Region code
- country. Country

Author(s)

Marc Hoeglinger and Ben Jann

References

https://ideas.repec.org/p/bss/wpaper/17.html

Plagiarism	*An Experimental Survey Measuring Plagiarism Using the Crosswise Model*

Description

A dataset containing the responses to sensitive questions about plagiarism and other attributes of 812 students. The crosswise model (CM) and direct questioning (DQ) were utilized to gather the data. Each row holds the response to one question for one student. The variables are as follows:

Usage

```
data(Plagiarism)
```

Format

A data frame in long format with 812 rows and 24 variables

Details

- id. Identification code of the student
- question. Which question was asked (1 and 3: Partial Plagiarism, 2 and 4: Severe Plagiarism)
- response. Binary randomized response
- gender. Gender of the student (0: male, 1: female)
- age. Age in years
- nationality. Nationality of the student (0: German or Swiss, 1: other)
- no_papers. Number of papers
- uni. Location of data collection (1: ETH Zurich, 2: LMU Munich, 3: University Leipzig)
- course. Course in which the data was collected
- Aspired_Degree. Aspired degree of the student
- Semester. semesters enrolled
- ur_none. Used resources: none
- ur_books. Used resources: books
- ur_art. Used resources: articles
- ur_int. Used resources: internet
- ur_fsp. Used resources: fellow students' papers
- ur_other. Used resources: other
- preading. Proofreading
- gradesf. Satisfaction with grades
- pp. Plagiarism indicator (0: Severe Plagiarism, 1: Partial Plagiarism)
- RR. Randomized Response indicator (0: DQ, 1: Crosswise)
- RRp1. Randomized Response parameter p1
- RRp2. Randomized Response parameter p2
- RRmodel. Randomized Response Model

Author(s)

Ben Jann and Laurence Brandenberger

References

doi: 10.7892/boris.51190

---

plot.RRglm   *Plot diagnostics for a RRglm object*

Description

Six plots (selectable by which) are currently available: (1) a plot of estimated population prevalence per RR model, (2) a plot of estimated population prevalence per protection level, (3) a plot of ungrouped residuals against fitted response probability, (4) a plot of grouped (on covariates) residuals against fitted response probability, (5) a plot of grouped Hosmer-Lemeshow residuals against fitted response probability, and (6) a Normal Q-Q plot of grouped (on covariates) residuals. By default, plots 1, 3, 4 and 6 are provided.

Usage

```
## S3 method for class 'RRglm'
plot(
  x,
  which = c(1, 3, 4, 6),
  type = c("deviance", "pearson"),
  ngroups = 10,
  ...
)
```

Arguments

x	an object of class RRglm.
which	if a subset of the plots is required, specify a subset of the numbers 1:6 (default: 1, 3, 4, 6).
type	the type of residuals which should be used to be used for plots 3, 4 and 6. The alternatives are: "deviance" (default) and "pearson".
ngroups	the number of groups to compute the Hosmer-Lemeshow residuals for (default: 10).
...	further arguments passed to or from other methods.

Examples

```
out <- RRglm(response ~ Gender + RR + pp + age, link="RRlink.logit", RRmodel=RRmodel,
        p1=RRp1, p2=RRp2, data=Plagiarism, etastart=rep(0.01, nrow(Plagiarism)))
plot(out, which = 1:6, type = "deviance", ngroups = 50)
```

---

plot.RRglmerMod   *Plot diagnostics for a RRglmerMod object*

Description

Five plots (selectable by which) are currently available: (1) a plot of estimated population preva-
lence per RR model, (2) a plot of estimated population prevalence per protection level, (3) a plot of
random effects and their conditional variance (95 (4) a plot of conditional pearson residuals against
fitted randomized response probability, and (5) a plot of unconditional pearson residuals against
fitted randomized response probability. By default, plots 1, 3, 4 and 5 are provided.

Usage

```
## S3 method for class 'RRglmerMod'
plot(x, which = c(1, 3, 4, 5), ...)
```

Arguments

x	an object of class RRglmerMod.
which	if a subset of the plots is required, specify a subset of the numbers 1:5 (default: 1, 3, 4, 5).
...	further arguments passed to or from other methods.

Examples

```
out <- RRglmer(response ~ Gender + RR + pp + (1+pp|age), link="RRlink.logit", RRmodel=RRmodel,
         p1=RRp1, p2=RRp2, data=Plagiarism, na.action = ”na.omit”,
         etastart = rep(0.01, nrow(Plagiarism)),
         control = glmerControl(optimizer = ”Nelder_Mead”, tolPwrss = 1e-03), nAGQ = 1)
plot(out, which = 1:5)
```

---

print.RRglmGOF   *Print RRglmGOF values*

Description

Print RRglmGOF values

Usage

```
## S3 method for class 'RRglmGOF'
print(x, digits = 3, ...)
```

Arguments

x	an object of class RRglmGOF.
digits	minimal number of *significant digits* (default: 3).
...	further arguments passed to or from other methods.

---

print.summary.RRglm   *Print RRglm summary*

Description

Print RRglm summary

Usage

```
## S3 method for class 'summary.RRglm'
print(
  x,
  printPrevalence = TRUE,
  printPrevalencePerLevel = FALSE,
  printResiduals = FALSE,
  digits = 5,
  ...
)
```

Arguments

x	an object of class summary.RRglm.
printPrevalence	print estimated population prevalence per item and RR model (default: true).
printPrevalencePerLevel	print estimated population prevalence per item, RRmodel and protection level (default: false).
printResiduals	print deviance residuals (default: false).
digits	minimal number of *significant digits* (default: 5).
...	further arguments passed to or from other methods.

print.summary.RRglmerMod

Print RRglmer summary

Description

Print RRglmer summary

Usage

```
## S3 method for class 'summary.RRglmerMod'
print(
  x,
  printPrevalence = TRUE,
  printPrevalencePerLevel = FALSE,
  printResiduals = FALSE,
  digits = 5,
  ...
)
```

Arguments

x	an object of class summary.RRglmerMod.
printPrevalence	
	print estimated population prevalence per item and RR model (default: true).
printPrevalencePerLevel	
	print estimated population prevalence per item, RRmodel and protection level (default: false).
printResiduals	print conditional deviance residuals (default: false).
digits	minimal number of *significant digits* (default: 5).
...	further arguments passed to or from other methods.

residuals.RRglm *Accessing GLMMRR Fits for fixed-effect models*

Description

Compute residuals for RRglm objects. Extends residuals.glm with residuals for grouped binary Randomized Response data.

Usage

```
## S3 method for class 'RRglm'
residuals(
  object,
 type = c("deviance", "pearson", "working", "response", "partial", "deviance.grouped",
    "pearson.grouped", "hosmer-lemeshow"),
  ngroups = 10,
  ...
)
```

Arguments

object	an object of class RRglm.
type	the type of residuals which should be returned. The alternatives are: "deviance" (default), "pearson", "working", "response", "partial", "deviance.grouped", "pearson.grouped" and "hosmer-lemeshow".
ngroups	the number of groups if Hosmer-Lemeshow residuals are computed (default: 10).
...	further arguments passed to or from other methods.

Value

A vector of residuals.

See Also

[residuals.glm](#)

---

residuals.RRglmerMod *Accessing GLMMRR Fits for mixed-effect models*

Description

Compute residuals for RRglmer objects. Extends [residuals.glmResp](#) to access conditional and unconditional residuals for grouped binary Randomized Response data.

Usage

```
## S3 method for class 'RRglmerMod'
residuals(
  object,
  type = c("deviance", "pearson", "working", "response", "partial",
    "unconditional.response", "unconditional.pearson"),
  ...
)
```

Arguments

object	an object of class RRglmer.
type	the type of residuals which should be returned. The alternatives are: "deviance" (default), "pearson", "working", "response", "partial", "unconditional.response" and "unconditional.pearson".
...	further arguments passed to or from other methods.

Value

A vector of residuals.

See Also

[residuals.glmResp](#)

RRbinomial	*Binomial family adjusted for Randomized Response parameters.*

Description

The upper and lower limits for mu's depend on the Randomized Response parameters.

Usage

```
RRbinomial(link, c, d, ...)
```

Arguments

link	a specification for the model link function. Must be an object of class "link-glm".
c	a numeric vector containing the parameter c.
d	a numeric vector containing the parameter d.
...	other potential arguments to be passed to [binomial](#).

Value

A binomial family object.

See Also

[family](#)

RRglm	*Fitting Generalized Linear Models with binary Randomized Response data*

Description

Fit a generalized linear model (GLM) with binary Randomized Response data. Implemented as a wrapper for [glm](). Reference: Fox, J-P, Veen, D. and Klotzke, K. (2018). Generalized Linear Mixed Models for Randomized Responses. *Methodology*. https://doi.org/10.1027/1614-2241/a000153

Usage

```
RRglm(formula, link, item, RRmodel, p1, p2, data, na.action = "na.omit", ...)
```

Arguments

formula	a two-sided linear formula object describing the model to be fitted, with the response on the left of a ~ operator and the terms, separated by + operators, on the right.
link	a glm link function for binary outcomes. Must be a function name. Available options: "RRlink.logit", "RRlink.probit", "RRlink.cloglog" and "RRlink.cauchit"
item	optional item identifier for long-format data.
RRmodel	the Randomized Response model, defined per case. Available options: "DQ", "Warner", "Forced", "UQM", "Crosswise", "Triangular" and "Kuk"
p1	the Randomized Response parameter p1, defined per case. Must be 0 <= p1 <= 1.
p2	the Randomized Response parameter p2, defined per case. Must be 0 <= p2 <= 1.
data	a data frame containing the variables named in [formula]() as well as the Randomized Response model and parameters. If the required information cannot be found in the data frame, or if no data frame is given, then the variables are taken from the environment from which RRglm is called.
na.action	a function that indicates what should happen when the data contain NAs. The default action ([na.omit](), as given by getOption("na.action"))) strips any observations with any missing values in any variables.
...	other potential arguments to be passed to [glm]().

Value

An object of class RRglm. Extends the class glm with Randomize Response data.

See Also

[glm]()

Examples

```
# Fit the model with fixed effects for gender, RR, pp and age using the logit link function.
# The Randomized Response parameters p1, p2 and model
# are specified for each observation in the dataset.
out <- RRglm(response ~ Gender + RR + pp + age, link="RRlink.logit", RRmodel=RRmodel,
         p1=RRp1, p2=RRp2, data=Plagiarism, etastart=rep(0.01, nrow(Plagiarism)))
summary(out)
```

RRglmer	*Fitting Generalized Linear Mixed-Effects Models with binary Randomized Response data*

Description

Fit a generalized linear mixed-effects model (GLMM) with binary Randomized Response data. Both fixed effects and random effects are specified via the model formula. Randomize response parameters can be entered either as single values or as vectors. Implemented as a wrapper for [glmer](). Reference: Fox, J-P, Veen, D. and Klotzke, K. (2018). Generalized Linear Mixed Models for Randomized Responses. *Methodology.* https://doi.org/10.1027/1614-2241/a000153

Usage

```
RRglmer(
  formula,
  item,
  link,
  RRmodel,
  p1,
  p2,
  data,
  control = glmerControl(),
  na.action = "na.omit",
  ...
)
```

Arguments

formula	a two-sided linear formula object describing both the fixed-effects and fixed-effects part of the model, with the response on the left of a ~ operator and the terms, separated by + operators, on the right. Random-effects terms are distinguished by vertical bars ("|") separating expressions for design matrices from grouping factors.
item	optional item identifier for long-format data.
link	a glm link function for binary outcomes. Must be a function name. Available options: "RRlink.logit", "RRlink.probit", "RRlink.cloglog" and "RRlink.cauchit"
RRmodel	the Randomized Response model, defined per case. Available options: "DQ", "Warner", "Forced", "UQM", "Crosswise", "Triangular" and "Kuk"

p1	the Randomized Response parameter p1, defined per case. Must be 0 <= p1 <= 1.
p2	the Randomized Response parameter p2, defined per case. Must be 0 <= p2 <= 1.
data	a data frame containing the variables named in [formula](#) as well as the Randomized Response model and parameters. If the required information cannot be found in the data frame, or if no data frame is given, then the variables are taken from the environment from which RRglmer is called.
control	a list (of correct class, resulting from [lmerControl](#)() or [glmerControl](#)() respectively) containing control parameters, including the nonlinear optimizer to be used and parameters to be passed through to the nonlinear optimizer, see the *lmerControl documentation for details.
na.action	a function that indicates what should happen when the data contain NAs. The default action ([na.omit](#), as given by getOption("na.action"))) strips any observations with any missing values in any variables.
...	other potential arguments to be passed to [glmer](#).

Value

An object of class RRglmerMod. Extends the class glmerMod with Randomize Response data, for which many methods are available (e.g. methods(class="glmerMod")).

See Also

[lme4](#)

Examples

```
# Fit the model with fixed effects for gender, RR and pp
# and a random effect for age using the logit link function.
# The Randomized Response parameters p1, p2 and model
# are specified for each observation in the dataset.
out <- RRglmer(response ~ Gender + RR + pp + (1|age), link="RRlink.logit", RRmodel=RRmodel,
        p1=RRp1, p2=RRp2, data=Plagiarism, na.action = "na.omit",
        etastart = rep(0.01, nrow(Plagiarism)),
        control = glmerControl(optimizer = "Nelder_Mead", tolPwrss = 1e-03), nAGQ = 1)
summary(out)
```

RRglmGOF   *Goodness-of-fit statistics for binary Randomized Response data*

Description

Compute goodness-of-fit statistics for binary Randomized Response data. Pearson, Deviance and Hosmer-Lemeshow statistics are available.

Usage

```
RRglmGOF(
  RRglmOutput,
  doPearson = TRUE,
  doDeviance = TRUE,
  doHlemeshow = TRUE,
  hlemeshowGroups = 10,
  rm.na = TRUE
)
```

Arguments

RRglmOutput	a model fitted with the [RRglm](#) function.
doPearson	compute Pearson statistic.
doDeviance	compute Deviance statistic.
doHlemeshow	compute Hosmer-Lemeshow statistic.
hlemeshowGroups	
	number of groups to split the data into for the Hosmer-Lemeshow statistic (default: 10).
rm.na	remove cases with missing data.

Value

an option of class RRglmGOF.

Examples

```
out <- RRglm(response ~ Gender + RR + pp + age, link="RRlink.logit", RRmodel=RRmodel,
         p1=RRp1, p2=RRp2, data=Plagiarism, etastart=rep(0.01, nrow(Plagiarism)))
RRglmGOF(RRglmOutput = out, doPearson = TRUE, doDeviance = TRUE, doHlemeshow = TRUE)
```

RRlink.cauchit *Cauchit link function with Randomized Response parameters.*

Description

Cauchit link function with Randomized Response parameters.

Usage

```
RRlink.cauchit(c, d)
```

Arguments

c	a numeric vector containing the parameter c.
d	a numeric vector containing the parameter d.

Value

RR link function.

RRlink.cloglog   *Log-Log link function with Randomized Response parameters.*

Description

Log-Log link function with Randomized Response parameters.

Usage

```
RRlink.cloglog(c, d)
```

Arguments

c   a numeric vector containing the parameter c.

d   a numeric vector containing the parameter d.

Value

RR link function.

RRlink.logit   *Logit link function with Randomized Response parameters.*

Description

Logit link function with Randomized Response parameters.

Usage

```
RRlink.logit(c, d)
```

Arguments

c   a numeric vector containing the parameter c.

d   a numeric vector containing the parameter d.

Value

RR link function.

RRlink.probit   *Probit link function with Randomized Response parameters.*

Description

Probit link function with Randomized Response parameters.

Usage

```
RRlink.probit(c, d)
```

Arguments

c	a numeric vector containing the parameter c.
d	a numeric vector containing the parameter d.

Value

RR link function.

---

summary.RRglm   *Summarizing GLMMRR fits for fixed-effect models*

Description

Summarizing GLMMRR fits for fixed-effect models

Usage

```
## S3 method for class 'RRglm'
summary(object, p1p2.digits = 2, ...)
```

Arguments

object	an object of class RRglm.
p1p2.digits	number of digits for aggregating data based on the level of protection (default: 2).
...	further arguments passed to or from other methods.

Value

An object of class summary.RRglm. Extends the class `summary.glm` with Randomize Response data.

summary.RRglmerMod   *Summarizing GLMMRR fits for fixed-effect models*

Description

Summarizing GLMMRR fits for fixed-effect models

Usage

```
## S3 method for class 'RRglmerMod'
summary(object, p1p2.digits = 2, ...)
```

Arguments

object   an object of class RRglm.

p1p2.digits   number of digits for aggregating data based on the level of protection (default:
                2).

...   further arguments passed to or from other methods.

Value

An object of class summary.RRglmerMod. Extends the class summary.glmerMod with Randomize
Response data.

Index