Package 'GLMcat'

March 7, 2021

Title Generalized Linear Models for Categorical Responses

Version 0.2.1

Description In statistical modeling, there is a wide variety of regression models for categorical dependent variables (nominal or ordinal data); yet, there is no software embracing all these models together in a uniform and generalized format. Following the methodology proposed by Peyhardi, Trottier, and Guédon (2015) <doi:10.1093/biomet/asv042>, we introduce 'GLMcat', an R package to estimate generalized linear models implemented under the unified specification (r, F, Z). Where r represents the ratio of probabilities (reference, cumulative, adjacent, or sequential), F the cumulative distribution function for the linkage, and Z, the design matrix.

License GPL-3

Encoding UTF-8

Depends R (>= 2.10)

LazyData true

RoxygenNote 7.0.2

LinkingTo Rcpp, BH, RcppEigen

Imports Rcpp, stats

Suggests knitr, rmarkdown, testthat, dplyr, ggplot2, gridExtra, gtools, tidyr

VignetteBuilder knitr

NeedsCompilation yes

Author Lorena León [aut, cre], Jean Peyhardi [aut], Catherine Trottier [aut]

Maintainer Lorena León <ylorenaleonv@gmail.com>

Repository CRAN

Date/Publication 2021-03-06 23:50:39 UTC

R topics documented:

`coef.glmcat` *Model coefficients*

Description

Extract model coefficients from a glmcat object.

Usage

```
## S3 method for class 'glmcat'
coef(object, na.rm = FALSE, ...)
```

Arguments

`object`	a GLMcat model.
`na.rm`	TRUE for NA coefficients to be removed, default is FALSE.
`...`	other arguments.

Examples

```
data(DisturbedDreams)
mod1 <- GLMcat(
  formula = Level ~ Age,
  ref_category = "Very.severe",
  data = DisturbedDreams, distribution = "logistic"
)
coef(mod1)
```

`Discrete_CM` *Family of models for Discrete Choice*

Description

Discrete choice model: Requires data in long form. For each individual (or decision maker), there are multiple observations (rows), one for each of the alternatives the individual could have chosen. We call the group of observations for an individual a "case". Each case represents a single statistical observation although it comprises multiple observations.

Usage

```
Discrete_CM(
  formula = NaN,
  case_id = "a",
  alternatives = "a",
  reference = NaN,
  alternative_specific = NA_character_,
  data = c(1, NA, NaN, Inf, -Inf),
  distribution = "a",
  freedom_degrees = 1,
  intercept = "standard"
)
```

Arguments

`formula`	a symbolic description of the model to be fit. An expression of the form y ~ predictors is interpreted as a specification that the response y is modelled by a linear predictor specified symbolically by model. A particularity for the formula is that for the case-specific variables, the user can define a specific effect for a category.
`case_id`	a string with the name of the column that identifies each case.
`alternatives`	a string with the name of the column that identifies the vector of alternatives the individual could have chosen.
`reference`	a string indicating the reference category
`alternative_specific`	a character vector with the name of the explanatory variables that are different for each case, these are the alternative specific variables. By default, the case specific variables are the explanatory variables that are not identify in here, but that are part of the formula.
`data`	a dataframe object in R, with the dependent variable as factor.
`distribution`	a string indicating the F distribution, options are: logistic, normal, cauchit, student (any df), gompertz, gumbel.
`freedom_degrees`	an optional scalar to indicate the degrees of freedom for the Student distribution.
`intercept`	if "conditional" then the design will be equivalent to the conditional logit model

Note

For these models it is not allowed to exclude the intercept.

Examples

```
library(GLMcat)
data(TravelChoice)
Discrete_CM(formula = choice ~ hinc + gc + invt,
case_id = "indv",alternatives = "mode",reference = "air",
data = TravelChoice,  alternative_specific = c("gc", "invt"),
distribution = "logistic")
```

`DisturbedDreams` *Severity of disturbed dreams*

Description

$Boy's disturbed dreams$ benchmark dataset drawn from a study that cross-classified boys by their age, and the severity (not severe, severe 1, severe 2, very severe) of their disturbed dreams (Maxwell, 1961).

Usage

```
data(DisturbedDreams)
```

Format

A dataframe containing :

Age Individuals age

Level Severity level: Not.severe, Severe.1, Severe.2, Very.severe.

References

Maxwell, A.E. (1961) *Analysing qualitative data*, Methuen London, 73.

Examples

```
data(DisturbedDreams)
```

GLMcat *Family of models for categorical responses (reference, adjacent, sequential and cumulative ratio)*

Description

Family of models for categorical responses (reference, adjacent, sequential and cumulative ratio)

Usage

```
GLMcat(
  formula,
  data,
  ratio = "reference",
  distribution = "logistic",
  proportional = NA_character_,
  categories_order = NA_character_,
  ref_category = NA_character_,
  freedom_degrees = NaN,
  threshold = "standard",
  beta_init = NA_real_
)
```

Arguments

formula	a symbolic description of the model to be fit. An expression of the form y ~ predictors is interpreted as a specification that the response y is modelled by a linear predictor specified symbolically by model.
data	a dataframe object in R, with the dependent variable as factor.
ratio	a string indicating the F distribution, options are: reference, adjacent, cumulative and sequential. Default value is reference.
distribution	a string indicating the F distribution, options are: logistic, normal, cauchit, student (any df), gompertz, gumbel.
proportional	a character vector indicating the name of the variables with a proportional effect. If variable is categorical, specify the name and the level of the variable as a string "namelevel".
categories_order	a character vector indicating the incremental order of the categories: c("a", "b", "c"); a<b<c. Alphabetical order is assumed by default. Order is relevant for adjacent, cumulative and sequential ratio.
ref_category	a string indicating the reference category. Proper option for models with reference ratio.
freedom_degrees	an optional scalar to indicate the degrees of freedom for the Student distribution.
threshold	restriction to impose on the thresholds, options are: standard, equidistant or symmetric (Valid only for the cumulative ratio).
beta_init	optional beta initialization vector.

Examples

```
data(DisturbedDreams)
ref_log_com <- GLMcat(formula = Level ~ Age, data = DisturbedDreams,
    ref_category = "Very.severe",
    distribution = "logistic", ratio = "reference")
```

`logLik.glmcat` *LogLikelihood glmcat models*

Description

Extract LogLikelihood for GLMcat models.

Usage

```
## S3 method for class 'glmcat'
logLik(object, ...)
```

Arguments

`object`	a GLMcat model.
`...`	other arguments.

Examples

```
data(DisturbedDreams)
mod1 <- GLMcat(
  formula = Level ~ Age,
  categories_order = c("Not.severe", "Severe.1", "Severe.2", "Very.severe"),
  data = DisturbedDreams, distribution = "logistic"
)
logLik(mod1)
```

`nobs_glmcat` *Number of observations in a glmcat model*

Description

Extract the number of observations from a GLMcat model.

Usage

```
nobs_glmcat(object, ...)
```

Arguments

object	a GLMcat model.
...	other arguments.

Examples

```
data(DisturbedDreams)
mod1 <- GLMcat(
  formula = Level ~ Age,
  categories_order = c("Not.severe", "Severe.1", "Severe.2", "Very.severe"),
  data = DisturbedDreams, distribution = "logistic"
)
nobs_glmcat(mod1)
```

---

predict_glmcat *GLMcat model predictions*

Description

GLMcat model predictions

Usage

```
predict_glmcat(model_object = NaN, data, type = "prob")
```

Arguments

model_object	a GLMcat model
data	a data frame with the predictor variables used in the GLMcat model.
type	The type of prediction to obtain. "prob" gives probabilities, "cum.prob" gives cumulative probabilities and "linear.predict" gives the linear predictor.

Examples

```
data(DisturbedDreams)
mod1 <- GLMcat(formula = Level ~ Age,
data = DisturbedDreams, distribution = "logistic")
predict_glmcat(mod1, data = DisturbedDreams[1:5, ], type = "prob")
```

summary.glmcat *Summary of models*

Description

summary method for GLMcat objects.

Usage

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

Arguments

object	a GLMcat model
...	additional arguments affecting the summary produced.

TravelChoice *Travel Mode Choice*

Description

The data set contains 210 observations on mode choice for travel between Sydney and Melbourne, Australia.

Usage

```
data(TravelChoice)
```

Format

A dataframe containing :

indv Id of the individual

mode available options: air, train, bus or car

choice a logical vector indicating as TRUE the transportation mode chosen by the traveler As category-specific variables:

invt travel time in vehicle

gc generalized cost measure

ttme terminal waiting time for plane, train and bus; 0 for car

invc in vehicle cost As case-specific variables:

hinc household income

psize traveling group size in mode chosen

Source

Download from on-line (18/09/2020) complements to Greene, W.H. (2011) Econometric Analysis, Prentice Hall, 7th Edition http://people.stern.nyu.edu/wgreene/Text/Edition7/TableF18-2.csv, Table F18-2.

References

Greene, W.H. and D. Hensher (1997) *Multinomial logit and discrete choice models in* Greene, W. H. (1997) *LIMDEP version 7.0 user's manual revised*, Plainview, New York econometric software, Inc .

Examples

```
data(TravelChoice)
```

Index