Package ‘GLMcat’

May 18, 2021

Title Generalized Linear Models for Categorical Responses

Version 0.2.3

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 cdf function for the linkage, and Z, the design matrix.

License GPL-3

Encoding UTF-8

Depends R (>= 2.10)

LazyData true

RoxygenNote 7.1.1

LinkingTo Rcpp, BH, RcppEigen

Imports Rcpp, stats

Suggests knitr, markdown, 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 2021-05-18 11:30:06 UTC

R topics documented:

    coef.glmcat .................................................. 2
coef.glmcat

control.glmcat .................................................. 3
Discrete_CM ...................................................... 3
DisturbedDreams ................................................. 5
GLMcat ............................................................. 5
logLik.glmcat ..................................................... 7
nobs_glmcat ...................................................... 7
noncentralt.glmcat ............................................... 8
predict_glmcat ................................................... 8
step_glmcat ....................................................... 9
student.glmcat .................................................. 10
summary.glmcat ................................................ 10
TravelChoice ..................................................... 11

Index 12

<table>
<thead>
<tr>
<th>coef.glmcat</th>
<th>Model coefficients</th>
</tr>
</thead>
</table>

### Description

Extract model coefficients from a glmcat object.

### Usage

```r
## 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

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

Set control parameters for GLMcat models.

Usage

control.glmcat(maxit = 25, epsilon = 1e-06, beta_init = NA)

Arguments

maxit the maximum number of the Fisher’s Scoreg Algorithm iterations. Defaults to 25.
epsilon a double to change update the convergence criterion of GLMcat models.
beta_init an appropriate sized vector for the initial iteration of the algorithm.

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),
  cdf = NaN,
  intercept = "standard",
  normalization = 1,
  control = NaN
)
Arguments

formula a symbolic description of the model to be fit. An expression of the form y \sim 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 (in a long format) object in R, with the dependent variable as factor.

cdf cdf: a string indicating the F cdf, options are: logistic, normal, cauchy, student (any df), noncentralt, gompertz, gumbel and laplace.

df: an integer with the degrees of freedom of the ‘cdf’

mu: an integer with the mu parameter of the ‘cdf’

intercept if "conditional" then the design will be equivalent to the conditional logit model

normalization the quantile to use for the normalization of the estimated coefficients where the logistic distribution is used as the base cumulative distribution function.

control maxit: the maximum number of iterations for the Fisher scoring algorithm.

epsilon: a double with to fix the epsilon value

beta_init: an appropriate sized vector for the initial iteration of the algorithm

Details

Family of models for Discrete Choice

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"),
cdf = "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**

```r
data(DisturbedDreams)
```

**Format**

A dataframe containing:

- **Age**: Individuals age
- **Level**: Severity level: Not.severe, Severe.1, Severe.2, Very.severe.

**References**


**Examples**

```r
data(DisturbedDreams)
```

---

**GLMcat**  
*GLMcat*

**Description**

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

**Usage**

```r
GLMcat(
  formula,
  data,
  ratio = "reference",
  cdf = NaN,
  parallel = NA_character_,
  categories_order = NA_character_,
  ref_category = NA_character_,
  threshold = "standard",
  control = NaN,
  normalization = 1
)
```
## Arguments

- **formula**: a symbolic description of the model to be fit. An expression of the form y \sim 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 cdf, options are: reference, adjacent, cumulative and sequential. Default value is reference.

- **cdf**: a string indicating the F cdf, options are: logistic, normal, cauchy, student (any df), noncentralt, gompertz, gumbel and laplace.

- **df**: an integer with the degrees of freedom of the 'cdf'

- **mu**: an integer with the mu parameter of the 'cdf'

- **parallel**: a character vector indicating the name of the variables with a parallel 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.

- **threshold**: restriction to impose on the thresholds, options are: standard, equidistant or symmetric (Valid only for the cumulative ratio).

- **control**: maxit: the maximum number of iterations for the Fisher scoring algorithm.

- **epsilon**: a double to change update the convergence criterion of GLMcat models.

- **beta_init**: an appropiate sized vector for the initial iteration of the algorithm.

- **normalization**: the quantile to use for the normalization of the estimated coefficients where the logistic distribution is used as the base cumulative distribution function.

## Details

Families of models for categorical responses

## Examples

```r
data(DisturbedDreams)
ref_log_com <- GLMcat(formula = Level ~ Age, data = DisturbedDreams,
                       ref_category = "Very.severe",
                       cdf = "logistic", ratio = "reference")
```
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, cdf = "logistic"
)
logLik(mod1)

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, cdf = "logistic"
)
nobs_glmcat(mod1)

---

noncentralt.glmcat | Noncentral t cdf for glmcat models

Description

Noncentral t cdf

Usage

noncentralt.glmcat(df = 7, mu = 0)

Arguments

df degrees_freedom
mu non centrality parameter

---

predict_glmcat | predict.glmcat

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.

Details

Prediction based on GLMcat models
Examples

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

---

**step_glmcat**

*Stepwise for glmcat models*

**Description**

Stepwise based on the AIC

**Usage**

```r
step_glmcat(
  object,
data,
scope,
direction = c("both", "backward", "forward"),
trace = 1,
steps = 1000
)
```

**Arguments**

- **object**: a GLMcat model.
- **data**: the same dataset used for the model specified in object.
- **scope**: defines the range of models examined in the stepwise search (same as in the step function of the stats package). This should be either a single formula, or a list containing components upper and lower, both formulae.
- **direction**: the mode of the stepwise search.
- **trace**: to print the process information.
- **steps**: the maximum number of steps.
student.glmcat

Description

Student distribution

Usage

student.glmcat(df = 7)

Arguments

df degrees_freedom

summary.glmcat

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.
**Description**

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

**Usage**

```r
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
- 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
- Case-specific variables:
  - `hinc` household income
  - `psize` traveling group size in mode chosen

**Source**


**References**


**Examples**

```r
data(TravelChoice)
```
Index

* datasets
  DisturbedDreams, 5
  TravelChoice, 11

coef.glmcat, 2
control.glmcat, 3

Discrete_CM, 3
DisturbedDreams, 5

GLMcat, 5

logLik.glmcat, 7

nobs_glmcat, 7
noncentral.t.glmcat, 8

predict_glmcat, 8

step_glmcat, 9
student.glmcat, 10
summary.glmcat, 10

TravelChoice, 11