Package ‘GLMcat’

September 23, 2021

Title  Generalized Linear Models for Categorical Responses

Version  0.2.4

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, stringr

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

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  2021-09-23 15:10:02 UTC

R topics documented:

anova.glmcat  ................................................................. 2
**anova.glmcat**

An analysis of deviance table for one fitted glmcat model object.

### Description

Compute an analysis of deviance table for one fitted glmcat model object.

### Usage

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

### Arguments

- **object**: an object of class "glmcat".
- **...**: additional arguments.
**coef.glmcat**

*Model coefficients of a fitted glmcat model object*

**Description**

Returns the coefficient estimates of the fitted glmcat model object.

**Usage**

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

**Arguments**

- `object`: an fitted object of class glmcat.
- `na.rm`: TRUE for NA coefficients to be removed, default is FALSE.
- `...`: additional arguments affecting the `coef` method.

---

**confint.glmcat**

*Confidence intervals for parameters of a fitted glmcat model object*

**Description**

Computes confidence intervals from a fitted glmcat model object for all the parameters.

**Usage**

```r
## S3 method for class 'glmcat'
confint(object, parm, level, ...)
```

**Arguments**

- `object`: an fitted object of class glmcat.
- `parm`: a numeric or character vector indicating which regression coefficients should be displayed
- `level`: the confidence level.
- `...`: other parameters.
Discrete Choice Models

Description

Fit discrete choice models which require 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. A group of observations of the same individual is a "case". Remark that each case represents a single statistical observation although it comprises multiple observations.

Usage

```r
discrete_cm(
    formula,
    case_id,
    alternatives,
    reference,
    alternative_specific = NA,
    data,
    cdf = list(),
    intercept = "standard",
    normalization = 1,
    control = list()
)
```
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**: The inverse distribution function to be used as part of the link function. If the distribution has no parameters to specify then it should be entered as a string indicating the name, e.g., \texttt{cdf = “normal”}, the default value is \texttt{cdf = “logistic”}. If there are parameters to specify then a list must be entered, so far this would only be the case for Student’s distribution which would be \texttt{list(“student”,df=2)}, and for the non-central distribution of student, \texttt{list(“noncentralt”,df=2,\mu=1)}.
- **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**: \texttt{maxit}: the maximum number of iterations for the Fisher scoring algorithm. \texttt{epsilon}: a double with to fix the epsilon value \texttt{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

```r
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)
```

---

### extractAIC.glmcat  Extract AIC from a fitted glmcat model object

**Description**

Method to compute the (generalized) Akaike An Information Criterion for a fitted object of class `glmcat`.

**Usage**

```r
## S3 method for class 'glmcat'
extractAIC(fit, ...)
```

**Arguments**

- **fit** an fitted object of class `glmcat`.
- **...** further arguments (currently unused in base R).
Description

Estimate generalized linear models implemented under the unified specification (ratio, cdf, Z) where ratio represents the ratio of probabilities (reference, cumulative, adjacent, or sequential), cdf the cumulative distribution function for the linkage, and Z the design matrix which must be specified through the parallel and the threshold arguments.

Usage

\[
\text{glmcat}(
\text{formula}, \\
\text{data}, \\
\text{ratio} = \text{c("reference", "cumulative", "sequential", "adjacent")}, \\
\text{cdf} = \text{list()}, \\
\text{parallel} = \text{NA}, \\
\text{categories\_order} = \text{NA}, \\
\text{ref\_category} = \text{NA}, \\
\text{threshold} = \text{c("standard", "symmetric", "equidistant")}, \\
\text{control} = \text{list()}, \\
\text{normalization} = 1, \\
\ldots
\)
\]

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 ratio (equivalently to the family) options are: reference, adjacent, cumulative and sequential. Default value is reference.

cdf: The inverse distribution function to be used as part of the link function. If the distribution has no parameters to specify then it should be entered as a string indicating the name, e.g., cdf = "normal", the default value is cdf = "logistic". If there are parameters to specify then a list must be entered, so far this would only be the case for Student’s distribution which would be list("student",df=2), and for the non-central distribution of student, list("noncentralt",df=2,\mu=1).

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 appropriate 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.

... additional arguments.

Details

Fitting models for categorical responses

References


Examples

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

logLik(glmcat)

Log-likelihood of a fitted glmcat model object

Description

Extract Log-likelihood of a fitted glmcat model object.

Usage

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

Arguments

object an fitted object of class glmcat.
... additional arguments affecting the loglik.
**nobs.glmcat**

Number of observations of a fitted glmcat model object

---

**Description**

Extract the number of observations of the fitted glmcat model object.

**Usage**

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

**Arguments**

- **object**: an fitted object of class glmcat.
- **...**: additional arguments affecting the nobs method.

---

**plot.glmcat**

Plot method for a fitted codeglmcat model object

---

**Description**

plot of the log-likelihood profile for a fitted glmcat model object.

**Usage**

```r
## S3 method for class 'glmcat'
plot(x, ...)
```

**Arguments**

- **x**: an object of class glmcat.
- **...**: additional arguments.
predict.glmcat

Predict method for a fitted glmcat model object

Description

Obtains predictions of a fitted glmcat model object.

Usage

## S3 method for class 'glmcat'
predict(object, newdata, type, ...)

Arguments

object a fitted object of class glmcat.
newdata optionally, a data frame in which to look for the variables involved in the model. If omitted, the fitted linear predictors are used.
type the type of prediction required. The default is "prob" which gives the probabilities, the other option is "linear.predictor" which gives predictions on the scale of the linear predictor.
... further arguments. The default is "prob" which gives the probabilities, the other option is "linear.predictor" which gives predictions on the scale of the linear predictor.

print.anova.glmcat

Printing Anova for glmcat model fits

Description

print.anova method for GLMcat objects.

Usage

## S3 method for class 'anova.glmcat'
print(x, digits = max(getOption("digits") - 2, 3), ...)

Arguments

x an object of class "glmcat".
digits the number of digits in the printed table.
... additional arguments affecting the summary produced.
print.glmcat

Print method for a fitted codeglmcat model object

Description

print method for a fitted glmcat model object.

Usage

## S3 method for class 'glmcat'
print(x, ...)

Arguments

x an object of class glmcat.
... additional arguments.

print.summary.glmcat

Printing a fitted glmcat model object

Description

print.summary method for GLMcat objects.

Usage

## S3 method for class 'summary.glmcat'
print(x, digits = max(3,getOption("digits") - 3), ...)

Arguments

x an object of class "glmcat".
digits the number of digits in the printed table.
... additional arguments affecting the summary produced.
**Stepwise for a glmcat model object**

**Description**
Stepwise for a glmcat model object based on the AIC.

**Usage**

```r
## S3 method for class 'glmcat'
step(object, scope, direction, trace, steps)
```

**Arguments**
- `object`: an fitted object of class glmcat.
- `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.

**Summary method for a fitted glmcat model object**

**Description**
Summary method for a fitted glmcat model object.

**Usage**

```r
## S3 method for class 'glmcat'
summary(object, normalized = FALSE, correlation = FALSE, ...)
```

**Arguments**
- `object`: an fitted object of class glmcat.
- `normalized`: if normalized is TRUE summary method yields the normalized coefficients.
- `correlation`: TRUE to print the Correlation Matrix.
- `...`: additional arguments affecting the summary produced.
**terms.glmcat**

**Terms of a fitted glmcat model object**

**Description**

Returns the terms of a fitted glmcat model object.

**Usage**

```r
## S3 method for class 'glmcat'
terms(x, ...)
```

**Arguments**

- `x`: an object of class `glmcat`.
- `...`: additional arguments.

**TravelChoice**

**Travel Mode Choice**

**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 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


References


Examples

data(TravelChoice)

---

vcov.glmcat

Variance-Covariance Matrix for a fitted glmcat model object

Description

Returns the variance-covariance matrix of the main parameters of a fitted glmcat model object.

Usage

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

Arguments

object an object of class glmcat.
... additional arguments.
Index

* datasets
  DisturbedDreams, 6
  TravelChoice, 13

anova.glmcat, 2
coef.glmcat, 3
confint.glmcat, 3
control_glmcat, 4
discrete_cm, 4
DisturbedDreams, 6
extractAIC.glmcat, 6
glmcat, 7
logLik.glmcat, 8
nobs.glmcat, 9
plot.glmcat, 9
predict.glmcat, 10
print.anova.glmcat, 10
print.glmcat, 11
print.summary.glmcat, 11
step.glmcat, 12
summary.glmcat, 12
terms.glmcat, 13
TravelChoice, 13
vcov.glmcat, 14