Package ‘EffectStars’

October 22, 2019

Type Package

Title Visualization of Categorical Response Models

Version 1.9-1

Date 2019-10-22

Depends VGAM

Author Gunther Schauberger

Maintainer Gunther Schauberger <gunther.schauberger@tum.de>

Description

Notice: The package EffectStars2 provides a more up-to-date implementation of effect stars! EffectStars provides functions to visualize regression models with categorical response as proposed by Tutz and Schauberger (2013) <doi:10.1080/10618600.2012.701379>. The effects of the variables are plotted with star plots in order to allow for an optical impression of the fitted model.

License GPL-2

LazyLoad yes

NeedsCompilation no

Repository CRAN

Date/Publication 2019-10-22 09:40:05 UTC

R topics documented:

`alligator` *Alligator Food*

Description

The data describe the food choice of alligators, they originate from a study of the Florida Game and Fresh Water Commission.

Usage

```
data(alligator)
```

Format

A data frame with 219 observations on the following 4 variables.

`Food` Food type with levels `bird`, `fish`, `invert`, `other` and `rep`

`Size` Size of the alligator with levels `<2.3` and `>2.3`

`Gender` Gender with levels `female` and `male`

`Lake` Name of the lake with levels `George`, `Hancock`, `Oklawaha` and `Trafford`

Source

http://www.stat.ufl.edu/~aa/cda/sas/sas.html

References

Agresti (2002): *Categorical Data Analysis*, Wiley.

Examples

```
## Not run:
data(alligator)
star.nominal(Food ~ Size + Lake + Gender, data = alligator, nlines = 2)

## End(Not run)
```

BEPS — *British Election Panel Study*

Description

These data are drawn from the 1997-2001 British Election Panel Study (BEPS).

Usage

```
data(BEPS)
```

Format

A data frame with 1525 observations on the following 10 variables.

- `Europe` An 11-point scale that measures respondents' attitudes toward European integration. High scores represent eurosceptic sentiment
- `Leader_Cons` Assessment of the Conservative leader Hague, 1 to 5
- `Leader_Labour` Assessment of the Labour leader Blair, 1 to 5
- `Leader_Liberals` Assessment of the Liberals leader Kennedy, 1 to 5
- `Vote` Party Choice with levels `Conservative`, `Labour` and `Liberal Democrat`
- `Age` Age in years
- `Gender` Gender with levels `female` and `male`
- `Political_Knowledge` Knowledge of parties' positions on European integration, 0 to 3
- `National_Economy` Assessment of current national economic conditions, 1 to 5
- `Household` Assessment of current household economic conditions, 1 to 5

Source

R package carData: `BEPS`

References

British Election Panel Study (BEPS)
J. Fox and R. Andersen (2006): *Effect displays for multinomial and proportional-odds logit models*. Sociological Methodology 36, 225–255

Examples

```
## Not run:
data(BEPS)

BEPS$Europe<-scale(BEPS$Europe)
BEPS$Age<-scale(BEPS$Age)
BEPS$Leader_Labour<-BEPS$Leader_Labour-BEPS$Leader_Cons
BEPS$Leader<-BEPS$Leader_Labour
```

```
BEPS$Leader_Liberals<-BEPS$Leader_Liberals-BEPS$Leader_Cons

star.nominal(Vote ~ Age + Household + National_Economy + Household + Leader +
Europe + Political_Knowledge + Gender, data = BEPS,
xij = list(Leader~Leader_Labour+Leader_Liberals), catstar = FALSE, symmetric = FALSE)

## End(Not run)
```

`coffee` *Coffee Brands*

Description

The data frame is part of a long-term panel about the choice of coffee brands in 2111 households. The explanatory variables either refer to the household as a whole or to the head of the household.

Usage

```
data(coffee)
```

Format

A data frame with 2111 observations on the following 8 variables.

`Education` Educational level with levels no `Highschool` and `Highschool`

`PriceSensitivity` Price sensitivity with levels `not sensitive` and `sensitive`

`Income` Income with levels `< 2499` and `>= 2500`

`SocialLevel` Social level with levels `high` and `low`

`Age` Age with levels `< 49` and `>= 50`

`Brand` Coffee Brand with levels `Jacobs`, `JacobsSpecial`, `Aldi`, `AldiSpecial`, `Eduscho`, `EduschoSpecial`, `Tchibo`, `TchiboSpecial` and `Others`

`Amount` Amount of packs with levels `1` and `>= 2`

`Persons` Number of persons in household

References

Gesellschaft für Konsumforschung (GfK)

Examples

```
## Not run:
data(coffee)

star.nominal(Brand ~ Amount + Age + SocialLevel + Income + Persons +
                 PriceSensitivity + Education, coffee, cex.cat = 0.5, cex.labels = 0.8)

## End(Not run)
```

EffectStars *Visualization of Categorical Response Models*

Description

The package EffectStars2 provides a more up-to-date implementation of effect stars!

The package provides functions that visualize categorical regression models.
Included models are the multinomial logit model, the sequential logit model and the cumulative logit model.
The exponentials of the effects of the predictors are plotted as star plots showing the strengths of the effects.
In addition p-values for the effect of predictors are given.
Various data sets and examples are provided.
The plots should in general be exported to file formats like pdf, ps or png to recieve the optimal display. Plotting in R devices may not provide the optimal results.

For further details see `star.nominal`, `star.sequential` and `star.cumulative`.

Author(s)

Gunther Schauberger
`<gunther.schauberger@tum.de>`
https://www.sg.tum.de/epidemiologie/team/schauberger/

References

Tutz, G. and Schauberger, G. (2012): *Visualization of Categorical Response Models - from Data Glyphs to Parameter Glyphs*, Journal of Computational and Graphical Statistics 22(1), 156-177.

Gerhard Tutz (2012): *Regression for Categorical Data*, Cambridge University Press

See Also

`star.nominal`, `star.sequential`, `star.cumulative`

election *Election Data*

Description

The data set contains data from the German Longitudinal Election Study. The Response Categories refer to the five dominant parties in Germany. The explanatory variables refer to the declarations of single voters.

Usage

```
data(election)
```

Format

A data frame with 816 observations on the following 30 variables.

`Age` Standardized age of the voter

`AgeOrig` Unstandardized age of the voter

`Partychoice` Party Choice with levels `CDU`, `SPD`, `FDP`, `Greens` and `Left Party`

`Gender` Gender with levels `female` and `male`

`West` Regional provenance (West-Germany or East-Germany) with levels `east` and `west`

`Union` Member of a Union with levels `no member` and `member`

`Highschool` Educational level with levels `no highschool` and `highschool`

`Unemployment` Unemployment with levels `not unemployed` and `unemployed`

`Pol.Interest` Political Interest with levels `very interested` and `less interested`

`Democracy` Satisfaction with the functioning of democracy with levels `satisfied` and `not satisfied`

`Religion` Religion with levels `evangelical`, `catholic` and `other religion`

`Social_CDU` Difference in attitude towards the socioeconomic dimension of politics between respondent and CDU

`Social_SPD` Difference in attitude towards the socioeconomic dimension of politics between respondent and SPD

`Social_FDP` Difference in attitude towards the socioeconomic dimension of politics between respondent and FDP

`Social_Greens` Difference in attitude towards the socioeconomic dimension of politics between respondent and the Greens

`Social_Left` Difference in attitude towards the socioeconomic dimension of politics between respondent and the Left party

`Immigration_CDU` Difference in attitude towards immigration of foreigners between respondent and CDU

`Immigration_SPD` Difference in attitude towards immigration of foreigners between respondent and SPD

`Immigration_FDP` Difference in attitude towards immigration of foreigners between respondent and FDP

`Immigration_Greens` Difference in attitude towards immigration of foreigners between respondent and the Greens

`Immigration_Left` Difference in attitude towards immigration of foreigners between respondent and the Left party

`Nuclear_CDU` Difference in attitude towards nuclear energy between respondent and CDU

`Nuclear_SPD` Difference in attitude towards nuclear energy between respondent and SPD

`Nuclear_FDP` Difference in attitude towards nuclear energy between respondent and FDP

`Nuclear_Greens` Difference in attitude towards nuclear energy between respondent and the Greens

`Nuclear_Left` Difference in attitude towards nuclear energy between respondent and the Left party

`Left_Right_CDU` Difference in attitude towards the positioning on a political left-right scale between respondent and CDU

`Left_Right_SPD` Difference in attitude towards the positioning on a political left-right scale between respondent and SPD

`Left_Right_FDP` Difference in attitude towards the positioning on a political left-right scale between respondent and FDP

`Left_Right_Greens` Difference in attitude towards the positioning on a political left-right scale between respondent and the Greens

`Left_Right_Left` Difference in attitude towards the positioning on a political left-right scale between respondent and the Left party

References

German Longitudinal Election Study (GLES)

Examples

```
## Not run:
data(election)

# simple multinomial logit model
star.nominal(Partychoice ~ Age + Religion + Democracy + Pol.Interest +
                 Unemployment + Highschool + Union + West + Gender, election)

# Use effect coding for the categorical predictor religion
star.nominal(Partychoice ~ Age + Religion + Democracy + Pol.Interest +
                 Unemployment + Highschool + Union + West + Gender, election,
                 pred.coding = "effect")

# Use reference category "FDP" instead of symmetric side constraints
star.nominal(Partychoice ~ Age + Religion + Democracy + Pol.Interest +
                 Unemployment + Highschool + Union + West + Gender, election,
                 refLevel = 3, symmetric = FALSE)

# Use category-specific covariates, subtract values for reference
# category CDU
election[,13:16] <- election[,13:16] - election[,12]
election[,18:21] <- election[,18:21] - election[,17]
election[,23:26] <- election[,23:26] - election[,22]
election[,28:31] <- election[,28:31] - election[,27]

election$Social <- election$Social_SPD
election$Immigration <- election$Immigration_SPD
election$Nuclear <- election$Nuclear_SPD
election$Left_Right <- election$Left_Right_SPD

star.nominal(Partychoice ~ Social + Immigration + Nuclear + Left_Right + Age +
Religion + Democracy + Pol.Interest + Unemployment + Highschool + Union + West +
```

```
Gender, data = election,
xij = list(Social ~ Social_SPD + Social_FDP + Social_Greens + Social_Left,
Immigration ~ Immigration_SPD + Immigration_FDP + Immigration_Greens + Immigration_Left,
Nuclear ~ Nuclear_SPD + Nuclear_FDP + Nuclear_Greens + Nuclear_Left,
Left_Right ~ Left_Right_SPD + Left_Right_FDP + Left_Right_Greens + Left_Right_Left),
symmetric = FALSE)

## End(Not run)
```

`insolvency` *Insolvency data*

Description

The data set originates from the Munich founder study. The data were collected on business founders who registered their new companies at the local chambers of commerce in Munich and surrounding administrative districts. The focus was on survival of firms measured in 7 categories, the first six represent failure in intervals of six months, the last category represents survival time beyond 36 months.

Usage

```
data(insolvency)
```

Format

A data frame with 1224 observations on the following 16 variables.

- `Insolvency` Survival of firms in ordered categories with levels `1` < `2` < `3` < `4` < `5` < `6` < `7`
- `Sector` Economic Sector with levels `industry`, `commerce` and `service industry`
- `Legal` Legal form with levels `small trade`, `one man business`, `GmBH` and `GbR,KG,OHG`
- `Location` Location with levels `residential area` and `business area`
- `New_Foundation` New Foundation or take-over with levels `new foundation` and `take-over`
- `Pecuniary_Reward` Pecuniary reward with levels `main` and `additional`
- `Seed_Capital` Seed capital with levels < `25000` and > `25000`
- `Equity_Capital` Equity capital with levels `no` and `yes`
- `Debt_Capital` Debt capital with levels `no` and `yes`
- `Market` Market with levels `local` and `national`
- `Clientele` Clientele with levels `wide spread` and `small`
- `Degree` Educational level with levels `no A-levels` and `A-Levels`
- `Gender` Gender with levels `female` and `male`
- `Experience` Professional experience with levels < `10 years` and > `10 years`
- `Employees` Number of employees with levels `0 or 1` and > `2`
- `Age` Age of the founder at formation of the company

Source

Münchner Gründer Studie

References

Brüderl, J. and Preisendörfer, P. and Ziegler, R. (1996): *Der Erfolg neugegründeter Betriebe: eine empirische Studie zu den Chancen und Risiken von Unternehmensgründungen*, Duncker & Humblot.

Examples

```
## Not run:
data(insolvency)


star.sequential(Insolvency ~ Sector + Legal + Pecuniary_Reward + Seed_Capital
+ Debt_Capital + Employees, insolvency, test.glob = FALSE, globcircle = TRUE, dist.x = 1.3)

star.cumulative(Insolvency ~ Sector + Employees, insolvency, select = 2:4)

## End(Not run)
```

PID *Party Identification*

Description

Subset of the 1996 American National Election Study.

Usage

```
data(election)
```

Format

A data frame with 944 observations on the following 6 variables.

`TVnews` Days in the past week spent watching news on TV

`PID` Party identification with levels `Democrat`, `Independent` and `Republican`

`Income` Income

`Education` Educational level with levels `low` (no college) and `high` (at least college)

`Age` Age in years

`Population` Population of respondent's location in 1000s of people

Source

R package faraway: `nes96`

Examples

```
## Not run:
data(PID)
PID$TVnews <- scale(PID$TVnews)
PID$Income <- scale(PID$Income)
PID$Age <- scale(PID$Age)
PID$Population <- scale(PID$Population)

star.nominal(PID ~ TVnews + Income + Population + Age + Education, data = PID)

## End(Not run)
```

---

plebiscite *Chilean Plebiscite*

Description

The data origin from a survey refering to the plebiscite in Chile 1988. The chilean people had to decide, wether Augusto Pinochet would remain president for another ten years (voting yes) or if there would be presidential elections in 1989 (voting no).

Usage

```
data(plebiscite)
```

Format

A data frame with 2431 observations on the following 7 variables.

`Gender` Gender with levels `female` and `male`

`Education` Educational level with levels `low` and `high`

`SantiagoCity` Respondent from Santiago City with levels `no` and `yes`

`Income` Monthly Income in Pesos

`Population` Population size of respondent's community

`Age` Age in years

`Vote` Response with levels `Abstention`, `No`, `Undecided` and `Yes`

Source

R package carData: `Chile`

References

Personal communication from FLACSO/Chile.
Fox, J. (2008): *Applied Regression Analysis and Generalized Linear Models*, Second Edition.

Examples

```
## Not run:
data(plebiscite)
plebiscite$Population <- scale(plebiscite$Population)
plebiscite$Age <- scale(plebiscite$Age)
plebiscite$Income <- scale(plebiscite$Income)

star.nominal(Vote ~ SantiagoCity + Population + Gender + Age + Education +
Income, data = plebiscite)

## End(Not run)
```

`star.cumulative` *Effect stars for cumulative logit models*

Description

The package EffectStars2 provides a more up-to-date implementation of effect stars!

The function computes and visualizes cumulative logit models. The computation is done with help of the package **VGAM**. The visualization is based on the function `stars` from the package **graphics**.

Usage

```
star.cumulative(formula, data, global = NULL, test.rel = TRUE, test.glob = FALSE,
    partial = FALSE, globcircle = FALSE, maxit = 100, scale = TRUE,
    nlines = NULL, select = NULL, dist.x = 1, dist.y = 1, dist.cov = 1,
    dist.cat = 1, xpd = TRUE, main = "", col.fill = "gray90",
    col.circle = "black", lwd.circle = 1, lty.circle = "longdash",
   col.global = "black", lwd.global = 1, lty.global = "dotdash", cex.labels = 1,
    cex.cat = 0.8, xlim = NULL, ylim = NULL)
```

Arguments

`formula`	An object of class “formula”. Formula for the cumulative logit model to be fitted and visualized.
`data`	An object of class “data.frame” containing the covariates used in `formula`.
`global`	Numeric vector to choose a subset of predictors to be included with global coefficients. Default is to include all coefficients category-specific. Numbers refer to total amount of predictors, including intercept and dummy variables.
`test.rel`	Provides a Likelihood-Ratio-Test to test the relevance of the explanatory covariates. The corresponding p-values will be printed as `p-rel`. `test.rel=FALSE` might save a lot of time. See also `Details`.
`test.glob`	Provides a Likelihood-Ratio-Test to test if a covariate has to be included as a category-specific covariate (in contrast to being global). The corresponding p-values will be printed as `p-global`. `test.glob=FALSE` and `globcircle=FALSE` might save a lot of time. See also `Details`.

partial	If partial=TRUE, partial proportional odds models with only one category-specific covariate are fitted. The resulting effects of the (sub)models are plotted. For further information see Details.
globcircle	If TRUE, additional circles that represent the global effects of the covariates are plotted. test.glob=FALSE and globcircle=FALSE might save a lot of time.
maxit	Maximal number of iterations to fit the cumulative logit model. See also vglm.control.
scale	If TRUE, the stars are scaled to equal maximal ray length.
nlines	If specified, nlines gives the number of lines in which the effect stars are plotted.
select	Numeric vector to choose only a subset of the stars to be plotted. Default is to plot all stars. Numbers refer to total amount of predictors, including intercept and dummy variables.
dist.x	Optional factor to increase/decrease distances between the centers of the stars on the x-axis. Values greater than 1 increase, values smaller than 1 decrease the distances.
dist.y	Optional factor to increase/decrease distances between the centers of the stars on the y-axis. Values greater than 1 increase, values smaller than 1 decrease the distances.
dist.cov	Optional factor to increase/decrease distances between the stars and the covariates labels above the stars. Values greater than 1 increase, values smaller than 1 decrease the distances.
dist.cat	Optional factor to increase/decrease distances between the stars and the category labels around the stars. Values greater than 1 increase, values smaller than 1 decrease the distances.
xpd	If FALSE, all plotting is clipped to the plot region, if TRUE, all plotting is clipped to the figure region, and if NA, all plotting is clipped to the device region. See also par.
main	An overall title for the plot. See also plot.
col.fill	Color of background of the circle. See also col in par.
col.circle	Color of margin of the circle. See also col in par.
lwd.circle	Line width of the circle. See also lwd in par.
lty.circle	Line type of the circle. See also lty in par.
col.global	Color of margin of the global effects circle. See also col in par. Ignored, if globcircle = FALSE.
lwd.global	Line width of the global effects circle. See also lwd in par. Ignored, if globcircle = FALSE.
lty.global	Line type of the global effects circle. See also lty in par. Ignored, if globcircle = FALSE.
cex.labels	Size of labels for covariates placed above the corresponding star. See also cex in par.
cex.cat	Size of labels for categories placed around the corresponding star. See also cex in par.
xlim	Optional specification of the x coordinates ranges. See also xlim in plot.window
ylim	Optional specification of the y coordinates ranges. See also ylim in plot.window

Details

The underlying models are fitted with the function `vglm` from the package **VGAM**. The family argument for `vglm` is `cumulative(parallel=FALSE)`.

The stars show the exponentials of the estimated coefficients. In cumulative logit models the exponential coefficients can be interpreted as odds. More precisely, the exponential $e^{\gamma_{rj}}, r = 1, \ldots, k-1$ represents the multiplicative effect of the covariate j on the cumulative odds $\frac{P(Y \leq r|x)}{P(Y > r|x)}$ if x_j increases by one unit.

In addition to the stars, we plot a cirlce that refers to the case where the coefficients of the corresponding star are zero. Therefore, the radii of these circles are always $exp(0) = 1$. If `scale=TRUE`, the stars are scaled so that they all have the same maximal ray length. In this case, the actual appearances of the circles differ, but they still refer to the no-effects case where all the coefficients are zero. Now the circles can be used to compare different stars based on their respective circles radii. The p-values beneath the covariate labels, which are given out if `test.rel=TRUE`, correspond to the distance between the circle and the star as a whole. They refer to a likelihood ratio test if all the coefficients from one covariate are zero (i.e. the variable is left out completely) and thus would lie exactly upon the cirlce.
The form of the circles can be modified by `col.circle`, `lwd.circle` and `lty.circle`.

By setting `globcircle=TRUE`, an addictional circle can be drawn. The radii now correspond to a model, where the respective covariate is not included category-specific but globally. Therefore, the distance between this circle and the star as a whole corresponds to the p-value p-global that is given if `test.glob=TRUE`.

Please note:

Regular fitting of cumulative logit models may fail because of the restrictions in the parameter space that have to be considered. If `partial=TRUE`, (sub)models with only one category-specific covariate, so-called partial proportional odds models, are fitted. Then at least estimates for every coefficient should be available. If `partial=TRUE`, the resulting effects of these (sub)models are plotted. It should be noted that in this case no coherent model is visualized. Also the p-values refer to the various submodels. For `partial=TRUE`, the p-values `p-rel` and `p-global` refer to tests of the corresponding partial proportial odds models against the proportional odds model.

It is strongly recommended to standardize metric covariates, display of effect stars can benefit greatly as in general differences between the coefficients are increased.

Value

P-values are only available if the corresponding option is set `TRUE`.

`odds`	Odds or exponential coefficients of the cumulative logit model
`coefficients`	Coefficients of the cumulative logit model
`se`	Standard errors of the coefficients
`p_rel`	P-values of Likelihood-Ratio-Tests for the relevance of the explanatory covariates

`p_global`	P-values of Likelihood-Ratio-Tests wether the covariates need to be included category-specific
`xlim`	`xlim` values that were automatically produced. May be helpfull if you want to specify your own `xlim`
`ylim`	`ylim` values that were automatically produced. May be helpfull if you want to specify your own `ylim`

Author(s)

Gunther Schauberger
<gunther.schauberger@tum.de>
https://www.sg.tum.de/epidemiologie/team/schauberger/

References

Tutz, G. and Schauberger, G. (2012): *Visualization of Categorical Response Models - from Data Glyphs to Parameter Glyphs*, Journal of Computational and Graphical Statistics 22(1), 156-177.

Gerhard Tutz (2012): *Regression for Categorical Data*, Cambridge University Press

See Also

`star.sequential`, `star.nominal`

Examples

```
## Not run:
data(insolvency)

star.cumulative(Insolvency ~ Sector + Employees, insolvency, select = 2:4)

## End(Not run)
```

`star.nominal` *Effect stars for multinomial logit models*

Description

The package EffectStars2 provides a more up-to-date implementation of effect stars!

The function computes and visualizes multinomial logit models. The computation is done with help of the package **VGAM**. The visualization is based on the function `stars` from the package **graphics**.

Usage

```
star.nominal(formula, data, xij = NULL, conf.int = FALSE, symmetric = TRUE,
   pred.coding = "reference", printpvalues = TRUE, test.rel = TRUE, refLevel = 1,
    maxit = 100, scale = TRUE, nlines = NULL, select = NULL, catstar = TRUE,
    dist.x = 1, dist.y = 1, dist.cov = 1, dist.cat = 1, xpd = TRUE, main = "",
    lwd.stars = 1, col.fill = "gray90", col.circle = "black", lwd.circle = 1,
    lty.circle = "longdash", lty.conf = "dotted", cex.labels = 1, cex.cat = 0.8,
    xlim = NULL, ylim = NULL)
```

Arguments

`formula` An object of class “formula”. Formula for the multinomial logit model to be fitted and visualized.

`data` An object of class “data.frame” containing the covariates used in `formula`.

`xij` An object of class list, used if category-specific covariates are to be inlcuded. Every element is a formula referring to one of the category-specific covariates. For details see help for `xij` in `vglm.control` and the details below.

`conf.int` If `TRUE`, confidence intervals are drawn.

`symmetric` Which side constraint for the coefficients in the multinomial logit model shall be used for the plot? Default `TRUE` uses symmetric side constraints, `FALSE` uses the reference category specified by `refLevel`. If category-specific covariates are specified using `xij`, automatically `symmetric = FALSE` is set. Symmetric side constraints are not possible in the case of category-specific covariates.

`pred.coding` Which coding for categorical predictors with more than two categories is to be used? Default `pred.coding="reference"` uses the first category as reference category, the alternative `pred.coding="effect"` uses effect coding equivalent to symmetric side constraints. For `pred.coding="effect"` a star for every category is plotted, for `pred.coding="reference"` no star for the reference category is plotted.

`printpvalues` If `TRUE`, p-values for the respective coefficients are printed besides the category labels. P-values are recieved by a Wald test.

`test.rel` Provides a Likelihood-Ratio-Test to test the relevance of the explanatory covariates. The corresponding p-values will be printed behind the covariates labels. `test.rel=FALSE` might save a lot of time.

`refLevel` Reference category for multinomial logit model. Ignored if `symmetric=TRUE`. See also `multinomial`.

`maxit` Maximal number of iterations to fit the multinomial logit model. See also `vglm.control`.

`scale` If `TRUE`, the stars are scaled to equal maximal ray length.

`nlines` If specified, `nlines` gives the number of lines in which the effect stars are plotted.

`select` Numeric vector to choose only a subset of the stars to be plotted. Default is to plot all stars. Numbers refer to total amount of predictors, including intercept and dummy variables.

catstar	A logical argument to specify if all category-specific effects in the model should be visualized with an additional star. Ignored if xij=NULL.
dist.x	Optional factor to increase/decrease distances between the centers of the stars on the x-axis. Values greater than 1 increase, values smaller than 1 decrease the distances.
dist.y	Optional factor to increase/decrease distances between the centers of the stars on the y-axis. Values greater than 1 increase, values smaller than 1 decrease the distances.
dist.cov	Optional factor to increase/decrease distances between the stars and the covariates labels above the stars. Values greater than 1 increase, values smaller than 1 decrease the distances.
dist.cat	Optional factor to increase/decrease distances between the stars and the category labels around the stars. Values greater than 1 increase, values smaller than 1 decrease the distances.
xpd	If FALSE, all plotting is clipped to the plot region, if TRUE, all plotting is clipped to the figure region, and if NA, all plotting is clipped to the device region. See also par.
main	An overall title for the plot. See also plot.
lwd.stars	Line width of the stars. See also lwd in par.
col.fill	Color of background of the circle. See also col in par.
col.circle	Color of margin of the circle. See also col in par.
lwd.circle	Line width of the circle. See also lwd in par.
lty.circle	Line type of the circle. See also lty in par.
lty.conf	Line type of confidence intervals. Ignored, if conf.int=FALSE. See also lty in par.
cex.labels	Size of labels for covariates placed above the corresponding star. See also cex in par.
cex.cat	Size of labels for categories placed around the corresponding star. See also cex in par.
xlim	Optional specification of the x coordinates ranges. See also xlim in plot.window
ylim	Optional specification of the y coordinates ranges. See also ylim in plot.window

Details

The underlying models are fitted with the function vglm from the package **VGAM**. The family argument for vglm is multinomial(parallel=FALSE).

The stars show the exponentials of the estimated coefficients. In multinomial logit models the exponential coefficients can be interpreted as odds. More precisely, for the model with symmetric side constraints, the exponential $e^{\gamma_{rj}}, r = 1, \ldots, k$ represents the multiplicative effect of the covariate j on the odds $\frac{P(Y=r|x)}{GM(x)}$ if x_j increases by one unit and $GM(x)$ is the median response. For the model with reference category k, the exponential $e^{\gamma_{rj}}, r = 1, \ldots, k-1$ represents the multiplicative effect of the covariate j on the odds $\frac{P(Y=r|x)}{P(Y=k|x)}$ if x_j increases by one unit.

In addition to the stars, we plot a cirlce that refers to the case where the coefficients of the corresponding star are zero. Therefore, the radii of these circles are always $exp(0) = 1$. If `scale=TRUE`, the stars are scaled so that they all have the same maximal ray length. In this case, the actual appearances of the circles differ, but they still refer to the no-effects case where all the coefficients are zero. Now the circles can be used to compare different stars based on their respective circles radii. The distances between the rays of a star and the cirlce correspond to the p-values that are printed beneath the category levels if `printpvalues=TRUE`. The closer a star ray lies to the no–effects circle, the more the p-value is increased.
The p-values beneath the covariate labels, which are given if `test.rel=TRUE`, correspond to the distance between the circle and the star as a whole. They refer to a likelihood ratio test if all the coefficients from one covariate are zero (i.e. the variable is left out completely) and thus would lie exactly upon the cirlce.
The appearance of the circles can be modified by `col.circle`, `lwd.circle` and `lty.circle`.

The argument `xij` is important because it has to be used to include category-specific covariates. If its default `xij=NULL` is kept, an ordinary multinomial logit model without category-specific covariates is fitted. If category-specific covariates are to be included, attention has to be paid to the exact usage of `xij`. Our `xij` argument is identical to the `xij` argument used in the embedded `vglm` function. For details see also `vglm.control`. The data are thought to be present in a wide format, i.e. a category-specific covariate consists of k columns. Before calling `star.nominal`, the values for the reference category (defined by `refLevel`) have to be subtracted from the values of the further categories. Additionally, the resulting variable for the first response category (but not the reference category) has to be duplicated. This duplicate should be denoted by an appropriate name for the category-specific variable, independent from the different response categories. It will be used as an assignment variable for the corresponding coefficient of the covariate and has to be included in to the `formula`. For every category-specific covariate, a formula has to be specified in the `xij` argument. On the left hand side of that formula, the assignment variable has to be placed. On the right hand side, the variables containing the differences from the values for the reference category are written. So the left hand side of the formula contains k-1 terms. The order of these terms has to be chosen according to the order of the response categories, ignoring the reference category. Examples for effect stars for models with category-specific covariates are recieved by typing `vignette("election")` or `vignette("plebiscite")`.

It is strongly recommended to standardize metric covariates, display of effect stars can benefit greatly as in general differences between the coefficients are increased.

Value

P-values are only available if the corresponding option is set `TRUE`.
`catspec` and `catspecse` are only available if `xij` is specified.

`odds`	Odds or exponential coefficients of the multinomial logit model
`coefficients`	Coefficients of the multinomial logit model
`se`	Standard errors of the coefficients
`pvalues`	P-values of Wald tests for the respective coefficients
`catspec`	Coefficients for the category-specific covariates

catspecse	Standard errors for the coefficients for the category-specific covariates
p_rel	P-values of Likelihood-Ratio-Tests for the relevance of the explanatory covariates
xlim	xlim values that were automatically produced. May be helpfull if you want to specify your own xlim
ylim	ylim values that were automatically produced. May be helpfull if you want to specify your own ylim

Author(s)

Gunther Schauberger
<gunther.schauberger@tum.de>
https://www.sg.tum.de/epidemiologie/team/schauberger/

References

Tutz, G. and Schauberger, G. (2012): *Visualization of Categorical Response Models - from Data Glyphs to Parameter Glyphs*, Journal of Computational and Graphical Statistics 22(1), 156-177.

Gerhard Tutz (2012): *Regression for Categorical Data*, Cambridge University Press

See Also

star.sequential, star.cumulative

Examples

```
## Not run:
data(election)

# simple multinomial logit model
star.nominal(Partychoice ~ Age + Religion + Democracy + Pol.Interest +
                 Unemployment + Highschool + Union + West + Gender, election)

# Use effect coding for the categorical predictor religion
star.nominal(Partychoice ~ Age + Religion + Democracy + Pol.Interest +
                 Unemployment + Highschool + Union + West + Gender, election,
                 pred.coding = "effect")

# Use reference category "FDP" instead of symmetric side constraints
star.nominal(Partychoice ~ Age + Religion + Democracy + Pol.Interest +
                 Unemployment + Highschool + Union + West + Gender, election,
                 refLevel = 3, symmetric = FALSE)

# Use category-specific covariates, subtract values for reference
# category CDU
election[,13:16] <- election[,13:16] - election[,12]
election[,18:21] <- election[,18:21] - election[,17]
election[,23:26] <- election[,23:26] - election[,22]
election[,28:31] <- election[,28:31] - election[,27]
```

```
election$Social <- election$Social_SPD
election$Immigration <- election$Immigration_SPD
election$Nuclear <- election$Nuclear_SPD
election$Left_Right <- election$Left_Right_SPD

star.nominal(Partychoice ~ Social + Immigration + Nuclear + Left_Right + Age +
Religion + Democracy + Pol.Interest + Unemployment + Highschool + Union + West +
Gender, data = election,
xij = list(Social ~ Social_SPD + Social_FDP + Social_Greens + Social_Left,
Immigration ~ Immigration_SPD + Immigration_FDP + Immigration_Greens + Immigration_Left,
Nuclear ~ Nuclear_SPD + Nuclear_FDP + Nuclear_Greens + Nuclear_Left,
Left_Right ~ Left_Right_SPD + Left_Right_FDP + Left_Right_Greens + Left_Right_Left),
symmetric = FALSE)

## End(Not run)
```

`star.sequential` *Effect stars for sequential logit models*

Description

The package EffectStars2 provides a more up-to-date implementation of effect stars!

The function computes and visualizes sequential logit models. The computation is done with help of the package **VGAM**. The visualization is based on the function `stars` from the package **graphics**.

Usage

```
star.sequential(formula, data, global = NULL, test.rel = TRUE, test.glob = FALSE,
    globcircle = FALSE, maxit = 100, scale = TRUE, nlines = NULL, select = NULL,
    dist.x = 1, dist.y = 1, dist.cov = 1, dist.cat = 1, xpd = TRUE, main = "",
    col.fill = "gray90", col.circle = "black", lwd.circle = 1,
    lty.circle = "longdash", col.global = "black", lwd.global = 1,
    lty.global = "dotdash", cex.labels = 1, cex.cat = 0.8, xlim = NULL,
    ylim = NULL)
```

Arguments

`formula`	An object of class “formula”. Formula for the sequential logit model to be fitted an visualized.
`data`	An object of class “data.frame” containing the covariates used in `formula`.
`global`	Numeric vector to choose a subset of predictors to be included with global coefficients. Default is to include all coefficients category-specific. Numbers refer to total amount of predictors, including intercept and dummy variables.
`test.rel`	Provides a Likelihood-Ratio-Test to test the relevance of the explanatory covariates. The corresponding p-values will be printed as `p-rel`. `test.rel=FALSE` might save a lot of time.

test.glob	Provides a Likelihood-Ratio-Test to test if a covariate has to be included as a category-specific covariate (in contrast to being global). The corresponding p-values will be printed as p-global. test.glob=FALSE and globcircle=FALSE might save a lot of time.
globcircle	If TRUE, additional circles that represent the global effects of the covariates are plotted. test.glob=FALSE and globcircle=FALSE might save a lot of time.
maxit	Maximal number of iterations to fit the sequential logit model. See also vglm.control.
scale	If TRUE, the stars are scaled to equal maximal ray length.
nlines	If specified, nlines gives the number of lines in which the effect stars are plotted.
select	Numeric vector to choose only a subset of the stars to be plotted. Default is to plot all stars. Numbers refer to total amount of predictors, including intercept and dummy variables.
dist.x	Optional factor to increase/decrease distances between the centers of the stars on the x-axis. Values greater than 1 increase, values smaller than 1 decrease the distances.
dist.y	Optional factor to increase/decrease distances between the centers of the stars on the y-axis. Values greater than 1 increase, values smaller than 1 decrease the distances.
dist.cov	Optional factor to increase/decrease distances between the stars and the covariates labels above the stars. Values greater than 1 increase, values smaller than 1 decrease the distances.
dist.cat	Optional factor to increase/decrease distances between the stars and the category labels around the stars. Values greater than 1 increase, values smaller than 1 decrease the distances.
xpd	If FALSE, all plotting is clipped to the plot region, if TRUE, all plotting is clipped to the figure region, and if NA, all plotting is clipped to the device region. See also par.
main	An overall title for the plot. See also plot.
col.fill	Color of background of the circle. See also col in par.
col.circle	Color of margin of the circle. See also col in par.
lwd.circle	Line width of the circle. See also lwd in par.
lty.circle	Line type of the circle. See also lty in par.
col.global	Color of margin of the global effects circle. See also col in par. Ignored, if globcircle = FALSE.
lwd.global	Line width of the global effects circle. See also lwd in par. Ignored, if globcircle = FALSE.
lty.global	Line type of the global effects circle. See also lty in par. Ignored, if globcircle = FALSE.
cex.labels	Size of labels for covariates placed above the corresponding star. See also cex in par.
cex.cat	Size of labels for categories placed around the corresponding star. See also cex in par.
xlim	Optional specification of the x coordinates ranges. See also xlim in plot.window
ylim	Optional specification of the y coordinates ranges. See also ylim in plot.window

Details

The underlying models are fitted with the function vglm from the package **VGAM**. The family argument for vglm is sratio(parallel=FALSE).

The stars show the exponentials of the estimated coefficients. In sequential logit models the exponential coefficients can be interpreted as odds. More precisely, the exponential $e^{\gamma_{rj}}, r = 1, \ldots, k-1$ represents the multiplicative effect of the covariate j on the continuation ratio odds $\frac{P(Y=r|x)}{P(Y>r|x)}$ if x_j increases by one unit.

In addition to the stars, we plot a cirlce that refers to the case where the coefficients of the corresponding star are zero. Therefore, the radii of these circles are always $exp(0) = 1$. If scale=TRUE, the stars are scaled so that they all have the same maximal ray length. In this case, the actual appearances of the circles differ, but they still refer to the no-effects case where all the coefficients are zero. Now the circles can be used to compare different stars based on their respective circles radii. The p-values beneath the covariate labels, which are given out if test.rel=TRUE, correspond to the distance between the circle and the star as a whole. They refer to a likelihood ratio test if all the coefficients from one covariate are zero (i.e. the variable is left out completely) and thus would lie exactly upon the cirlce.
The appearance of the circles can be modified by col.circle, lwd.circle and lty.circle.

By setting globcircle=TRUE, an addictional circle can be drawn. The radii now correspond to a model, where the respective covariate is not included category-specific but globally. Therefore, the distance between this circle and the star as a whole corresponds to the p-value p-global that is given if test.glob=TRUE.

It is strongly recommended to standardize metric covariates, display of effect stars can benefit greatly as in general differences between the coefficients are increased.

Value

P-values are only available if the corresponding option is set TRUE.

odds	Odds or exponential coefficients of the sequential logit model
coefficients	Coefficients of the sequential logit model
se	Standard errors of the coefficients
p_rel	P-values of Likelihood-Ratio-Tests for the relevance of the explanatory covariates
p_global	P-values of Likelihood-Ratio-Tests wether the covariates need to be included category-specific
xlim	xlim values that were automatically produced. May be helpfull if you want to specify your own xlim
ylim	ylim values that were automatically produced. May be helpfull if you want to specify your own ylim

Author(s)

Gunther Schauberger
<gunther.schauberger@tum.de>
https://www.sg.tum.de/epidemiologie/team/schauberger/

References

Tutz, G. and Schauberger, G. (2012): *Visualization of Categorical Response Models - from Data Glyphs to Parameter Glyphs*, Journal of Computational and Graphical Statistics 22(1), 156-177.

Gerhard Tutz (2012): *Regression for Categorical Data*, Cambridge University Press

See Also

star.nominal, star.cumulative

Examples

```
## Not run:
data(insolvency)

star.sequential(Insolvency ~ Sector + Legal + Pecuniary_Reward + Seed_Capital
+ Debt_Capital + Employees, insolvency, test.glob = FALSE, globcircle = TRUE, dist.x = 1.3)

## End(Not run)
```

---

womenlabour *Canadian Women's Labour-Force Participation*

Description

The data are from a 1977 survey of the Canadian population.

Usage

```
data(womenlabour)
```

Format

A data frame with 263 observations on the following 4 variables.

`Participation` Labour force participation with levels `fulltime`, `not.work` and `parttime`

`IncomeHusband` Husband's income in 1000 $

`Children` Presence od children in household with levels `absent` and `present`

`Region` Region with levels `Atlantic`, `BC`, `Ontario`, `Prairie` and `Quebec`

Source

R package carData: Womenlf

References

Social Change in Canada Project. York Institute for Social Research.
Fox, J. (2008): *Applied Regression Analysis and Generalized Linear Models*, Second Edition.

Examples

```
## Not run:
data(womenlabour)
womenlabour$IncomeHusband <- scale(womenlabour$IncomeHusband)

star.nominal(Participation ~ IncomeHusband + Children + Region, womenlabour)

## End(Not run)
```

Index