Package 'rioplot'

June 14, 2024

Type Package

Title Turn a Regression Model Inside Out

Version 1.1.1

Maintainer David Melamed <dmmelamed@gmail.com>

Description

Turns regression models inside out. Functions decompose variances and coefficients for various regression model types. Functions also visualize regression model objects using techniques developed in Schoon, Melamed, and Breiger (2024) <doi:10.1017/9781108887205>.

VignetteBuilder knitr

Depends R (>= 3.5.0), ggplot2, methods

Suggests dplyr, knitr, rmarkdown, ggrepel, MASS

License GPL-2 | GPL-3

Encoding UTF-8

LazyData true

NeedsCompilation no

Author David Melamed [aut, cre] (<https://orcid.org/0000-0002-8821-7698>),
Ronald L. Breiger [aut],
Eric W. Schoon [aut]

Repository CRAN

Date/Publication 2024-06-14 09:50:01 UTC

Contents

`Beckfield06` *Replication data for Beckfield (2006) as re-analyzed by Schoon, Melamed, and Breiger (2024)*

Description

Beckfield (2006) analyzed these data using fixed and random effects regression models. He showed that regional economic and political integregation is associated with increased economic inequality. Schoon, Melamed, and Breiger (2024) turned these models inside out and decomposed the model coefficients.

Usage

```
data("Beckfield06")
```

Format

A data frame with 48 observations on the following 9 variables.

`year` a numeric vector

`polint` a numeric vector

`ecoint` a numeric vector

`ecoints` a numeric vector

`gdp` a numeric vector

`trans` a numeric vector

`outflo` a numeric vector

`gini` a numeric vector

`countryid` a character vector

References

Beckfield, Jason. 2006. "European integration and income inequality."" American Sociological Review 71(6): 964-985. Schoon, Eric W., David Melamed, and Ronald L. Breiger. 2024. Regression Inside Out. NY: Cambridge University Press.

Examples

```
data(Beckfield06)
head(Beckfield06)
```

cosine *Compute the Cosine similarity between two points.*

Description

Given two points, the function computes the cosine similarity between them.

Usage

```
cosine(x,y)
```

Arguments

x	Point 1
y	Point 2

Value

The cosine similarity, ranging between -1 and +1.

Author(s)

Ronald L. Breiger, David Melamed and Eric Schoon

References

Schoon, Eric, David Melamed, and Ronald L. Breiger. 2023. Regression Inside Out. NY: Cambridge University Press.

Examples

```
data(Kenworthy99)
m1 <- lm(scale(dv) ~ scale(gdp) + scale(pov) + scale(tran) -1,data=Kenworthy99)
rp1 <- rio.plot(m1,include.int="no",r1=1:15)
cosine(rp1$row.dimensions[15,],rp1$row.dimensions[8,])
# cosine similarity between USA and Ireland

cosine(rp1$row.dimensions[15,],rp1$row.dimensions[14,])
# cosine similarity between USA and United Kingdom
```

`decompose.model` *Decompose the Results of a Regression Model by Cases*

Description

This function takes a regression model object and a vector of case assignments to groups (note, cases can be in their own group) and computes each cases' contribution to the overall regression coefficients.

Usage

```
decompose.model(m1,group.by=group.by,include.int="yes",model.type="OLS")
```

Arguments

`m1`	A regression model object. OLS, logistic, Poisson and negative binomial regression are supported.
`group.by`	A numeric vector denoting group membership. Should be the same length as the number of cases.
`include.int`	Whether the regression model included an intercept. Default is "yes."
`model.type`	Type of model to be decomposed. OLS via lm, logistic via glm ("logit"), Poisson via glm ("poisson"), and negative binomial via MASS ("nb") are supported.

Value

`decomp.coef`	Each case's or subset of cases' contribution to the estimated slope or regression coefficient.
`decomp.var`	Each case's or subset of cases' contribution to the variance of the estimated slope or regression coefficient.

Author(s)

David Melamed, Ronald L. Breiger, and Eric Schoon

References

Schoon, Eric, David Melamed, and Ronald L. Breiger. 2024. Regression Inside Out. NY: Cambridge University Press.

Examples

```
data(Kenworthy99)
m1 <- lm(scale(dv) ~ scale(gdp) + scale(pov) + scale(tran) -1,data=Kenworthy99)
decompose.model(m1,group.by=c("Liberal","Corp","Liberal",
"SocDem","SocDem","Corp","Corp","Corp","Corp","Corp","SocDem",
"SocDem","Liberal","Liberal","Liberal"),include.int="no")
```

GSS.2016 *Subset of data from the General Social Survey from 2016. Data were analyzed in Schoon, Melamed, and Breiger (2024).*

Description

Subset of data from the General Social Survey from 2016. Data were analyzed in Schoon, Melamed, and Breiger (2024). Full details on the variable selection and source information is available therein.

Usage

```
data("GSS.2016")
```

Format

A data frame with 2867 observations on the following 27 variables.

`sclass` a numeric vector
`fulltime` a numeric vector
`retired` a numeric vector
`hrsworked` a numeric vector
`occprestige` a numeric vector
`occprestige_partner` a numeric vector
`occprestige_mother` a numeric vector
`occprestige_father` a numeric vector
`children` a numeric vector
`age` a numeric vector
`educ` a numeric vector
`paeduc` a numeric vector
`maeduc` a numeric vector
`speduc` a numeric vector
`babs` a numeric vector
`female` a numeric vector
`white` a numeric vector
`black` a numeric vector
`other` a numeric vector
`income` a numeric vector
`republican` a numeric vector
`conservative` a numeric vector
`environment` a numeric vector
`helpblackpeople` a numeric vector
`science` a numeric vector
`govequalwealth` a numeric vector
`pclass` a numeric vector

References

Schoon, Eric W., David Melamed, and Ronald L. Breiger. 2024. Regression Inside Out. NY: Cambridge University Press.

Examples

```
data(GSS.2016)
head(GSS.2016)
```

GSS2018 *Subset of the General Social Survey analyzed by Schoon, Melamed, and Breiger (2024)*

Description

Subset of the General Social Survey analyzed by Schoon, Melamed, and Breiger (2024). Full details on the variable selection and source information is available therein.

Usage

```
data("GSS2018")
```

Format

A data frame with 558 observations on the following 7 variables.

`dog` a numeric vector

`race` a numeric vector

`sex` a numeric vector

`children` a numeric vector

`married` a numeric vector

`age` a numeric vector

`income` a numeric vector

References

Schoon, Eric W., David Melamed, and Ronald L. Breiger. 2024. Regression Inside Out. NY: Cambridge University Press.

Examples

```
data(GSS2018)
head(GSS2018)
```

Hilbe *Replication data for regression models with a count dependent variable.*

Description

Data analyzed by Hilbe (2011), and used here to illustrate model visualization and coefficient decomposition for count models.

Usage

```
data("Hilbe")
```

Format

A data frame with 601 observations on the following 9 variables.

`naffairs` a numeric vector

`avgmarr` a numeric vector

`hapavg` a numeric vector

`vryhap` a numeric vector

`smerel` a numeric vector

`vryrel` a numeric vector

`yrsmarr4` a numeric vector

`yrsmarr5` a numeric vector

`yrsmarr6` a numeric vector

Source

Hilbe, Joseph M., 2011. Negative binomial regression. NY: Cambridge University Press.

Examples

```
data(Hilbe)
head(Hilbe)
```

Kenworthy99 *Data to replicate OLS regression models reported in Kenworthy (1999).*

Description

Data to replicate OLS regression models reported in Kenworthy (1999). Data were analyzed in Schoon, Melamed, and Breiger (2024). Full details on the variable selection and source information is available therein.

Usage

```
data("Kenworthy99")
```

Format

A data frame with 15 observations on the following 6 variables.

`dv` a numeric vector

`gdp` a numeric vector

`pov` a numeric vector

`tran` a numeric vector

`ISO3` a character vector

`nation.long` a character vector

References

Kenworthy, Lane. 1999. "Do social-welfare policies reduce poverty? A cross-national assessment."" Social Forces 77(3): 1119-1139. Schoon, Eric W., David Melamed, and Ronald L. Breiger. 2024. Regression Inside Out. NY: Cambridge University Press.

Examples

```
data(Kenworthy99)
head(Kenworthy99)
```

`project.point` *Project point 1 onto the line (at 90 degress) running through point 2 and the origin (0,0).*

Description

Given two points, p1 and p2, this function identifies the point at which p1 is projected onto the line connecting p2 and the origin (0,0). The projection occurs at a right angle.

Usage

```
project.point(p1,p2)
```

Arguments

p1	First point, the one that is to be projected onto point 2.
p2	Second point, the one that is projected to the origin. This is the outcome or dependent variable in our book. See reference below.

Details

The output is just a single point. This is implemented as the point to which lines are drawn in many graphs.

Value

Two values which correspond to the x and y co-ordinates in the graph.

Author(s)

David Melamed, Ronald L. Breiger, and Eric Schoon

References

Schoon, Eric, David Melamed, and Ronald L. Breiger. 2024. Regression Inside Out. NY: Cambridge University Press.

Examples

```
data(Kenworthy99)
m1 <- lm(scale(dv) ~ scale(gdp) + scale(pov) + scale(tran) -1,data=Kenworthy99)
rp1 <- rio.plot(m1,include.int="no",r1=1:15)
project.point(as.numeric(rp1$col.dimensions[1,]),as.numeric(rp1$row.dimensions[1,]))
```

RaginData *Subset of replication data from Ragin and Fiss (2017).*

Description

Subset of replication data from Ragin and Fiss (2017). Data were analyzed in Schoon, Melamed, and Breiger (2024). Full details on the variable selection and source information is available therein.

Usage

```
data("RaginData")
```

Format

A data frame with 4185 observations on the following 10 variables.

`incrat` a numeric

`pinc` a numeric

`ped` a numeric

`resp_ed` a numeric

`afqt` a numeric

`kids` a numeric

`married` a numeric

`black` a numeric

`male` a numeric

`povd` a numeric

References

Ragin, Charles C. and Peer C. Fiss. 2017. Intersectional inequality: Race, class, test scores, and poverty. Chicago, IL: University of Chicago Press. Schoon, Eric W., David Melamed, and Ronald L. Breiger. 2024. Regression Inside Out. NY: Cambridge University Press.

Examples

```
data(RaginData)
head(RaginData)
```

rio.plot *Regression Inside Out: Plotting Regression Models*

Description

rio.plot is used to generate a reduced rank image of a regression model. The function computes row and column dimensions for both cases and variables, and generates an image of the model based on those scores.

Usage

```
rio.plot(m1,exclude.vars="no",r1="none",case.names="",col.names="no",
h.just=-.2,v.just=0,case.col="blue",var.name.col="black",
include.int="yes",group.cases=1,model.type="OLS")
```

Arguments

m1	a regression model object. Supported models include OLS, Logistic, Poisson, and Negative Binomial Regression.
exclude.vars	an optional numerical vector indicating variables from the model to exclude from the plot of the model.
r1	an optional numerical vector indicating cases to include in the plot. By default, all cases are excluded from the plot.
case.names	a character string of names to label the cases. Should be the same length as 'r1.'
col.names	whether to include the variable names in the plot. Default is "no"
h.just	horizontal justification in the plot. Default is -.2
v.just	vertical justification in the plot. Default is 0
case.col	if cases are added to the plot, this is their color. Default is "blue"
var.name.col	Color of the names of variables in the plot. Default is "black"
include.int	Whether the underlying model included a model intercept. Default is "yes"
group.cases	Whether to aggregate cases into clusters or subsets. If yes, provide a numeric vector of memberships. It will aggregate over them by summing.
model.type	The type of regression model. OLS is supported via the lm function. Logistic and Poisson regression are supported via the glm function. Negative Binomial regression is supported via the MASS package. Default is "OLS." For logistic regression, use "logit." For Poisson regression, use "poisson." For negative binomial regression, use "nb."

Details

The function take a regression model object (OLS, logistic, Poisson, or negative binomial) and computes the corresponding row (case) and column (variables) scores. The scores are part of the output, as is a ggplot object of the model.

Value

rio.plot returns several objects.

p1	a ggplot object of the model space, given the terms in the function
row.dimensions	the scores assigned to each case, or each subset of cases if they were aggregated using the 'group.cases' option. These are the co-ordinates in the plot.
col.dimensions	the scores assigned to each variable. These are the co-ordinates in the plot.
case.variances	each cases' contribution (or each subsets' contribution) to the variance of the estimated regression coefficient
U	The orthogonalized column space matrix from the Singular Value Decomposition of the predictor matrix and fitted values.
UUt	The orthogonalized column space matrix from the Singular Value Decomposition of the predictor matrix and fitted values, post-multiplied by its transpose.

Author(s)

David Melamed, Ronald L. Breiger, and Eric Schoon

References

Schoon, Eric, David Melamed, and Ronald L. Breiger. 2024. Regression Inside Out. NY: Cambridge University Press.

Examples

```
data(Kenworthy99)
m1 <- lm(scale(dv) ~ scale(gdp) + scale(pov) + scale(tran) -1,data=Kenworthy99)
rp1 <- rio.plot(m1,include.int="no")
names(rp1)
rp1$gg.obj
# rp1$gg.obj + ggplot2::scale_x_continuous(limits=c(-.55,1)) # useful option

rp2 <- rio.plot(m1,r1=1:15,case.names=paste(1:15),include.int="no")
rp2$gg.obj

Kenworthy99 <- data.frame(Kenworthy99,type=c("Liberal","Corp","Liberal",
"SocDem","SocDem","Corp","Corp","Corp","Corp","Corp","SocDem","SocDem",
"Liberal","Liberal","Liberal"))

rp3 <- rio.plot(m1,r1=1:15,group.cases=Kenworthy99$type,include.int="no")
rp3$gg.obj
# rp3$gg.obj + ggplot2::scale_x_continuous(limits=c(-1,20))
```

SchneiderAndMakszin06 *Subset of replication data from Schneider and Makszin (2014).*

Description

Subset of replication data from Schneider and Makszin (2014). Data were analyzed in Schoon, Melamed, and Breiger (2024). Full details on the variable selection and source information is available therein.

Usage

```
data("SchneiderAndMakszin06")
```

Format

A data frame with 30 observations on the following 36 variables.

`id` a character vector

`country` a character vector

`year` a numeric vector

`fde` a numeric vector

`fde_cilb` a numeric vector

`fde_ciub` a numeric vector

`wcoord` a numeric vector

`govint` a numeric vector

`ud` a numeric vector

`epl` a numeric vector

`socexp` a numeric vector

`eduexp` a numeric vector

`vet_un` a numeric vector

`lmexp` a numeric vector

`wagecov` a numeric vector

`vet_isced3` a numeric vector

`eduexp_pri` a numeric vector

`edu_terenr` a numeric vector

`vt_reg` a numeric vector

`vt_vap` a numeric vector

`compvote` a numeric vector

`fde2` a numeric vector

`low_fde_l` a numeric vector

high_fde_l a numeric vector

high_wc_l a numeric vector

high_int_l a numeric vector

high_ud_l a numeric vector

high_epl_l a numeric vector

high_socx_l a numeric vector

high_edux_l a numeric vector

high_lmx_l a numeric vector

high_vet_l a numeric vector

p1_y a numeric vector

p2_y a numeric vector

p3_y a numeric vector

sol_y a numeric vector

References

Schneider, Carsten Q., and Kristin Makszin. 2014. "Forms of welfare capitalism and education-based participatory inequality." Socio-Economic Review 12(2): 437-462. Schoon, Eric W., David Melamed, and Ronald L. Breiger. 2024. Regression Inside Out. NY: Cambridge University Press.

Examples

```
data(SchneiderAndMakszin06)
head(SchneiderAndMakszin06)
```

Wimmer_et_al_EPR *Subset of replication data from Wimmer, Cederman, and Min (2009).*

Description

Subset of replication data from Wimmer, Cederman, and Min (2009). Data were analyzed in Schoon, Melamed, and Breiger (2024). Full details on the variable selection and source information is available therein.

Usage

```
data("Wimmer_et_al_EPR")
```

Format

A data frame with 7908 observations on the following 80 variables.

`yearc` a numeric

`year` a numeric

`cowcode` a numeric

`country` a character

`gdpcap` a numeric

`gdpcapl` a numeric

`oilpc` a numeric

`oilpcl` a numeric

`popavg` a numeric

`lpopl` a numeric

`ethfrac` a numeric

`western` a numeric

`eeurop` a numeric

`lamerica` a numeric

`ssafrica` a numeric

`asia` a numeric

`nafrme` a numeric

`lmtnest` a numeric

`polity2` a numeric

`polity` a numeric

`anoc` a numeric

`anocl` a numeric

`democ` a numeric

`democl` a numeric

`regchg3` a numeric

`pimppast` a numeric

`groups` a numeric

`egipgrps` a numeric

`exclgrps` a numeric

`exclpop` a numeric

`lrexclpop` a numeric

`ttlpop` a numeric

`discpop` a numeric

`pwrlpop` a numeric

`olppop` a numeric

olpspop a numeric

jppop a numeric

sppop a numeric

dompop a numeric

monpop a numeric

maxexclpop a numeric

maxegippop a numeric

maxpop a numeric

newonset a numeric

newethonset a numeric

newhionset a numeric

newethhionset a numeric

onsetstatus a numeric

onsetstatus2 a numeric

actoraim a numeric

actoraim2 a numeric

ongoingwarl a numeric

ongoinghiwarl a numeric

newonset2 a numeric

newhionset2 a numeric

newethonset2 a numeric

warlfl a numeric

onsetfl a numeric

ethonsetfl a numeric

onsetfl2 a numeric

ethonsetfl2 a numeric

warstns2 a numeric

warstns1 a numeric

atwarnsl a numeric

npeaceyears a numeric

nspline1 a numeric

nspline2 a numeric

nspline3 a numeric

hpeaceyears a numeric

hspline1 a numeric

hspline2 a numeric

hspline3 a numeric

`fpeaceyears` a numeric

`fspline1` a numeric

`fspline2` a numeric

`fspline3` a numeric

`speaceyears` a numeric

`sspline1` a numeric

`sspline2` a numeric

`sspline3` a numeric

References

Wimmer, Andreas, Lars-Erik Cederman, and Brian Min. 2009. "Ethnic politics and armed conflict: A configurational analysis of a new global data set." American Sociological Review 74(2): 316-337.

Examples

```
data(Wimmer_et_al_EPR)
head(Wimmer_et_al_EPR)
```

Index