Package ‘rsq’

May 9, 2020

Title R-Squared and Related Measures

Version 2.0

Date 2020-5-8

Author Dabao Zhang

Maintainer Dabao Zhang <zhangdb@stat.purdue.edu>

Description
Calculate generalized R-squared, partial R-squared, and partial correlation coefficients for generalized linear (mixed) models (including quasi models with well defined variance functions).

Imports methods, stats, MASS, lme4, nlme, numDeriv, Matrix

Suggests

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2020-05-09 05:50:06 UTC

R topics documented:

`hcrabs` *Satellites of Female Horseshoe Crabs*

Description

Recorded are the numbers of male satellites, and other characteristics of 173 female horseshoe crabs.

Usage

```
data("hcrabs")
```

Format

A data frame with 173 observations on the following 5 variables.

- `color` the female crab's color, coded 1: light; 2: medium light; 3: medium; 4: medium dark; 5: dark. Not all of these colors appear.
- `spine` the female crab's spine condition, coded 1: both good; 2: one worn or broken; 3: both worn or broker.
- `width` the female crab's carapace width (cm).
- `num.satellites` the number of satellite males.
- `weight` the female crab's weight (kg).

Details

A nesting female horseshoe crab may have male crabs residing nearby, called satellites, besides the male crab residing in her nest. Brockmann (1996) investigated factors (including the female crab's color, spine condition, weight, and carapace width) which may influence the presence/obsence of satellite males. This data set has been discussed by Agresti (2002).

Author(s)

Dabao Zhang, Department of Statistics, Purdue University

Source

Agresti, A. (2012). *An Introduction to Categorical Data Analysis*, 3rd edition. Wiley: New Jersey.

References

Brockmann, H. J. (1996). Satellite male groups in horseshoe crabs. *Limulus polyphemus. Ethology*, 102: 1-21.

See Also

rsq, rsq.partial, pcor, simglm.

Examples

```
data(hcrabs)
summary(hcrabs)
head(hcrabs)

attach(hcrabs)
y <- ifelse(num.satellites>0,1,0)
bnfit <- glm(y~color+spine+width+weight,family=binomial)
rsq(bnfit)
rsq(bnfit,adj=TRUE)
rsq.partial(bnfit)

quasips <- glm(num.satellites~color+spine+width+weight,family=quasipoisson)
rsq(quasips)
rsq(quasips,adj=TRUE)
rsq.partial(quasips)
```

---

hschool *Attendance Behavior of High School Juniors*

Description

Recorded are the number of days of absence, gender, and two test scores of 316 high school juniors from two urban high schools.

Usage

```
data("hschool")
```

Format

A data frame with 316 observations on the following 5 variables.

`school` school of the two, coded 1 or 2;

`male` whether the student is male, coded 1: male; 0: female;

`math` the standardized test score for math;

`langarts` the standardized test score for language arts;

`daysabs` the number of days of absence.

Details

Some school administrators studied the attendance behavior of high school juniors at two schools. Predictors of the number of days of absence include gender of the student and standardized test scores in math and language arts. The original source of this data set is unknown.

Author(s)

Dabao Zhang, Department of Statistics, Purdue University

Source

http://www.ats.ucla.edu/stat/sas/dae/negbinreg.htm

See Also

rsq, rsq.partial, pcor, simglm.

Examples

```
data(hschool)
summary(hschool)
head(hschool)

require(MASS)
absfit <- glm.nb(daysabs~school+male+math+langarts,data=hschool)
summary(absfit)
rsq(absfit)
rsq(absfit,adj=TRUE)

rsq.partial(absfit)
```

---

lifetime   *Lifetimes in Two Different Environments.*

Description

There are 27 tests in each of the two environments.

Usage

```
data("lifetime")
```

Format

A data frame with 54 observations on the following 2 variables.

`time` the lifetime (x10).

`env` the environment of each test (kg/mm^2).

Details

This data set is discussed by Wang et al. (1992).

Author(s)

Dabao Zhang, Department of Statistics, Purdue University

Source

Wang, H., Ma, B., and Shi, J. (1992). Estimation of environmental factors for the inverse gaussian distribution. *Microelectron. Reliab.*, 32: 931-934.

See Also

`rsq`, `rsq.partial`, `pcor`, `simglm`.

Examples

```
data(lifetime)
summary(lifetime)
head(lifetime)

attach(lifetime)
igfit <- glm(time~env,family=inverse.gaussian)
rsq(igfit)
rsq(igfit,adj=TRUE)
```

---

pcor *Partial Correlation for Generalized Linear Models*

Description

Calculate the partial correlation for both linear and generalized linear models.

Usage

```
pcor(objF,objR=NULL,adj=FALSE,type=c('v','kl','sse','lr','n'))
```

Arguments

`objF`	an object of class "lm" or "glm", a result of a call to lm, glm, or glm.nb to fit the full model.
`objR`	an object of class "lm" or "glm", a result of a call to lm, glm, or glm.nb to fit the reduced model.
`adj`	logical; if TRUE, calculate the adjusted partial R^2.
`type`	the type of R-squared used: 'v' (default) – variance-function-based (Zhang, 2016), calling rsq.v; 'kl' – KL-divergence-based (Cameron and Windmeijer, 1997), calling rsq.kl; 'sse' – SSE-based (Efron, 1978), calling rsq.sse; 'lr' – likelihood-ratio-based (Maddala, 1983; Cox and Snell, 1989; Magee, 1990), calling rsq.lr; 'n' – corrected version of 'lr' (Nagelkerke, 1991), calling rsq.n.

Details

When the fitting object of the reduced model is not specified, the partial correlation of each covariate (excluding factor covariates with more than two levels) in the model will be calculated.

Value

The partial correlation coefficient is returned.

Author(s)

Dabao Zhang, Department of Statistics, Purdue University

References

Cameron, A. C. and Windmeijer, A. G. (1997) An R-squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics*, 77: 329-342.

Cox, D. R. and Snell, E. J. (1989) *The Analysis of Binary Data*, 2nd ed. London: Chapman and Hall.

Efron, B. (1978) Regression and ANOVA with zero-one data: measures of residual variation. *Journal of the American Statistical Association*, 73: 113-121.

Maddala, G. S. (1983) *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University.

Magee, L. (1990) R^2 measures based on Wald and likelihood ratio joint significance tests. *The American Statistician*, 44: 250-253.

Nagelkerke, N. J. D. (1991) A note on a general definition of the coefficient of determination. *Biometrika*, 78: 691-692.

Zhang, D. (2017). A coefficient of determination for generalized linear models. *The American Statistician*, 71(4): 310-316.

See Also

rsq, rsq.partial.

Examples

```
data(hcrabs)
attach(hcrabs)
y <- ifelse(num.satellites>0,1,0)
bnfit <- glm(y~color+spine+width+weight,family=binomial)
rsq.partial(bnfit)

bnfitr <- glm(y~color+weight,family=binomial)
rsq.partial(bnfit,bnfitr)

quasibn <- glm(y~color+spine+width+weight,family=quasibinomial)
rsq.partial(quasibn)

quasibnr <- glm(y~color+weight,family=binomial)
rsq.partial(quasibn,quasibnr)
```

rsq *R-Squared for Generalized Linear (Mixed) Models*

Description

Calculate the coefficient of determination, aka R^2, for both linear and generalized linear (mixed) models.

Usage

```
rsq(fitObj,adj=FALSE,type=c('v','kl','sse','lr','n'))
```

Arguments

fitObj	an object of class "lm", "glm", "merMod", "lmerMod", or "lme"; usually a result of a call to lm, glm, glm.nb, lmer or glmer or glmer.nb in lme4, or lme in nlme.
adj	logical; if TRUE, calculate the adjusted R^2.
type	the type of R-squared (only applicable for generalized linear models): 'v' (default) – variance-function-based (Zhang, 2016), calling rsq.v; 'kl' – KL-divergence-based (Cameron and Windmeijer, 1997), calling rsq.kl; 'sse' – SSE-based (Efron, 1978), calling rsq.sse; 'lr' – likelihood-ratio-based (Maddala, 1983; Cox and Snell, 1989; Magee, 1990), calling rsq.lr; 'n' – corrected version of 'lr' (Nagelkerke, 1991), calling rsq.n.

Details

Calculate the R-squared for (generalized) linear models. For (generalized) linear mixed models, there are three types of R^2 calculated on the basis of observed response values, estimates of fixed effects, and variance components, i.e., model-based R_M^2 (proportion of variation explained by the model in total, including both fixed-effects and random-efffects factors), fixed-effects R_F^2 (proportion of variation explained by the fixed-effects factors), and random-effects R_R^2 (proportion of variation explained by the random-effects factors).

Value

The R^2 or adjusted R^2. For (generalized) linear mixed models,

R_M^2	proportion of variation explained by the model in total, including both fixed-effects and random-efffects factors.
R_F^2	proportion of variation explained by the fixed-effects factors.
R_R^2	proportion of variation explained by the random-effects factors.

Author(s)

Dabao Zhang, Department of Statistics, Purdue University

References

Cameron, A. C. and Windmeijer, A. G. (1997) An R-squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics*, 77: 329-342.

Cox, D. R. and Snell, E. J. (1989) *The Analysis of Binary Data*, 2nd ed. London: Chapman and Hall.

Efron, B. (1978) Regression and ANOVA with zero-one data: measures of residual variation. *Journal of the American Statistical Association*, 73: 113-121.

Maddala, G. S. (1983) *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University.

Magee, L. (1990) R^2 measures based on Wald and likelihood ratio joint significance tests. *The American Statistician*, 44: 250-253.

Nagelkerke, N. J. D. (1991) A note on a general definition of the coefficient of determination. *Biometrika*, 78: 691-692.

Zhang, D. (2017). A coefficient of determination for generalized linear models. *The American Statistician*, 71(4): 310-316.

Zhang, D. (2020). Coefficients of determination for generalized linear mixed models. *Technical Report*, 20-01, Department of Statistics, Purdue University.

See Also

`rsq.partial`,`pcor`,`simglm`.

Examples

```
data(hcrabs)
attach(hcrabs)
y <- ifelse(num.satellites>0,1,0)
bnfit <- glm(y~color+spine+width+weight,family=binomial)
rsq(bnfit)
rsq(bnfit,adj=TRUE)

quasibn <- glm(y~color+spine+width+weight,family=quasibinomial)
rsq(quasibn)
rsq(quasibn,adj=TRUE)

psfit <- glm(num.satellites~color+spine+width+weight,family=poisson)
rsq(psfit)
rsq(psfit,adj=TRUE)

quasips <- glm(num.satellites~color+spine+width+weight,family=quasipoisson)
rsq(quasips)
rsq(quasips,adj=TRUE)

# Linear mixed models
require(lme4)
lmm1 <- lmer(Reaction~Days+(Days|Subject),data=sleepstudy)
rsq(lmm1)
rsq.lmm(lmm1)
```

```
# Generalized linear mixed models
data(cbpp)
glmm1 <- glmer(cbind(incidence,size-incidence)~period+(1|herd),data=cbpp,family=binomial)
rsq(glmm1)
```

`rsq.glmm` *R-Squared for Generalized Linear Mixed Models*

Description

Calculate the variance-function-based R-squared for generalized linear mixed models.

Usage

```
rsq.glmm(fitObj,adj=FALSE,DFUN=NULL)
```

Arguments

`fitObj`	an object of class "glmerMod", usually, a result of a call to glmer or glmer.nb in lme4.
`adj`	logical; if TRUE, calculate the adjusted R^2.
`DFUN`	function which calculates the directed length along the variance function, and should be specified for quasi models (with family=quasi()).

Details

There are three types of R^2 calculated on the basis of observed response values, estimates of fixed effects, and variance components, i.e., model-based R_M^2 (proportion of variation explained by the model in total, including both fixed-effects and random-efffects factors), fixed-effects R_F^2 (proportion of variation explained by the fixed-effects factors), and random-effects R_R^2 (proportion of variation explained by the random-effects factors).

Value

`R_M^2`	proportion of variation explained by the model in total, including both fixed-effects and random-efffects factors.
`R_F^2`	proportion of variation explained by the fixed-effects factors.
`R_R^2`	proportion of variation explained by the random-effects factors.

Author(s)

Dabao Zhang, Department of Statistics, Purdue University

References

Zhang, D. (2017). A coefficient of determination for generalized linear models. *The American Statistician*, 71(4): 310-316.

Zhang, D. (2020). Coefficients of determination for generalized linear mixed models. *Technical Report*, 20-01, Department of Statistics, Purdue University.

See Also

`vresidual`,`rsq`,`rsq.v`.

Examples

```
require(lme4)
data(cbpp)
glmm1 <- glmer(cbind(incidence,size-incidence)~period+(1|herd),data=cbpp,family=binomial)
rsq.glmm(glmm1)
rsq(glmm1)
```

---

rsq.kl *KL-Divergence-Based R-Squared*

Description

The Kullback-Leibler-divergence-based R^2 for generalized linear models.

Usage

```
rsq.kl(fitObj,adj=FALSE)
```

Arguments

`fitObj`	an object of class "lm" or "glm", usually, a result of a call to lm, glm, or glm.nb.
`adj`	logical; if TRUE, calculate the adjusted R^2.

Details

This version of R^2 was proposed by Cameron and Windmeijer (1997). It is extended to quasi models (Zhang, 2017) based on the quasi-likelihood function (McCullagh, 1983).

Value

The R^2 or adjusted R^2.

Author(s)

Dabao Zhang, Department of Statistics, Purdue University

References

Cameron, A. C. and Windmeijer, A. G. (1997) An R-squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics*, 77: 329-342.

McCullagh, P. (1983) Quasi-likelihood functions. *Annals of Statistics*, 11: 59-67.

See Also

rsq, rsq.partial, pcor.

Examples

```
data(hcrabs)
attach(hcrabs)
y <- ifelse(num.satellites>0,1,0)
bnfit <- glm(y~color+spine+width+weight,family=binomial)
rsq.kl(bnfit)
rsq.kl(bnfit,adj=TRUE)

psfit <- glm(num.satellites~color+spine+width+weight,family=poisson)
rsq.kl(psfit)
rsq.kl(psfit,adj=TRUE)

# Effectiveness of Bycycle Safety Helmets in Thompson et al. (1989)
y <- matrix(c(17,218,233,758),2,2)
x <- factor(c("yes","no"))
tbn <- glm(y~x,family=binomial)
rsq.kl(tbn)
rsq.kl(tbn,adj=TRUE)
```

---

rsq.lmm *R-Squared for Linear Mixed Models*

Description

Calculate the R-squared for linear mixed models.

Usage

```
rsq.lmm(fitObj,adj=FALSE)
```

Arguments

fitObj	an object of class "merMod" or "lmerMod" or "lme", usually, a result of a call to lmer in lme4, or lme in nlme.
adj	logical; if TRUE, calculate the adjusted R^2.

Details

There are three types of R^2 calculated on the basis of observed response values, estimates of fixed effects, and variance components, i.e., model-based R_M^2 (proportion of variation explained by the model in total, including both fixed-effects and random-efffects factors), fixed-effects R_F^2 (proportion of variation explained by the fixed-effects factors), and random-effects R_R^2 (proportion of variation explained by the random-effects factors).

Value

`R_M^2`	proportion of variation explained by the model in total, including both fixed-effects and random-efffects factors.
`R_F^2`	proportion of variation explained by the fixed-effects factors.
`R_R^2`	proportion of variation explained by the random-effects factors.

Author(s)

Dabao Zhang, Department of Statistics, Purdue University

References

Zhang, D. (2020). Coefficients of determination for generalized linear mixed models. *Technical Report*, 20-01, Department of Statistics, Purdue University.

See Also

`rsq`,`rsq.v`.

Examples

```
# lmer in lme4
require(lme4)
lmm1 <- lmer(Reaction~Days+(Days|Subject),data=sleepstudy)
rsq(lmm1)
rsq.lmm(lmm1)

# lme in nlme
require(nlme)
lmm2 <- lme(Reaction~Days,data=sleepstudy,random=~Days|Subject)
rsq(lmm2)
rsq.lmm(lmm2)
```

`rsq.lr` *Likelihood-Ratio-Based R-Squared*

Description

Calculate the likelihood-ratio-based R^2 for generalized linear models.

Usage

```
rsq.lr(fitObj,adj=FALSE)
```

Arguments

`fitObj`	an object of class "lm" or "glm", usually, a result of a call to lm, glm, or glm.nb.
`adj`	logical; if TRUE, calculate the adjusted R^2.

Details

Proposed by Maddala (1983), Cox and Snell (1989), and Magee (1990), this version of R^2 is defined with the likelihood ratio statistics, so it is not defined for quasi models. It reduces to the classical R^2 when the variance function is constant or linear.

Value

The R^2 or adjusted R^2.

Author(s)

Dabao Zhang, Department of Statistics, Purdue University

References

Cox, D. R. and Snell, E. J. (1989) *The Analysis of Binary Data*, 2nd ed. London: Chapman and Hall.

Maddala, G. S. (1983) *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University.

Magee, L. (1990) R^2 measures based on Wald and likelihood ratio joint significance tests. *The American Statistician*, 44: 250-253.

See Also

`rsq`, `rsq.partial`, `pcor`, `rsq.n`.

Examples

```
data(hcrabs)
attach(hcrabs)
y <- ifelse(num.satellites>0,1,0)
bnfit <- glm(y~color+spine+width+weight,family=binomial)
rsq.lr(bnfit)
rsq.lr(bnfit,adj=TRUE)

psfit <- glm(num.satellites~color+spine+width+weight,family=poisson)
rsq.lr(psfit)
rsq.lr(psfit,adj=TRUE)

# Effectiveness of Bycycle Safety Helmets in Thompson et al. (1989)
y <- matrix(c(17,218,233,758),2,2)
x <- factor(c("yes","no"))
tbn <- glm(y~x,family=binomial)
rsq.lr(tbn)
rsq.lr(tbn,adj=TRUE)
```

---

rsq.n *Corrected Likelihood-Ratio-Based R-Squared*

Description

Corrected likelihood-ratio-based R^2 for generalized linear models.

Usage

```
rsq.n(fitObj,adj=FALSE)
```

Arguments

`fitObj`	an object of class "lm" or "glm", usually, a result of a call to lm, glm, or glm.nb.
`adj`	logical; if TRUE, calculate the adjusted R^2.

Details

Nagelkerke (1991) proposed this version of R^2 to correct the likelihood-ratio-statistic-based one which was proposed by Maddala (1983), Cox and Snell (1989), and Magee (1990). This corrected generalization of R^2 cannot reduce to the classical R^2 in case of linear models. It is not defined for quasi models.

Value

The R^2 or adjusted R^2.

Author(s)

Dabao Zhang, Department of Statistics, Purdue University

References

Cox, D. R. and Snell, E. J. (1989) *The Analysis of Binary Data*, 2nd ed. London: Chapman and Hall.

Maddala, G. S. (1983) *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University.

Magee, L. (1990) R^2 measures based on Wald and likelihood ratio joint significance tests. *The American Statistician*, 44: 250-253.

Nagelkerke, N. J. D. (1991) A note on a general definition of the coefficient of determination. *Biometrika*, 78: 691-692.

See Also

rsq, rsq.partial, pcor, rsq.lr.

Examples

```
data(hcrabs)
attach(hcrabs)
y <- ifelse(num.satellites>0,1,0)
bnfit <- glm(y~color+spine+width+weight,family=binomial)
rsq.n(bnfit)
rsq.n(bnfit,adj=TRUE)

psfit <- glm(num.satellites~color+spine+width+weight,family=poisson)
rsq.n(psfit)
rsq.n(psfit,adj=TRUE)

# Effectiveness of Bycycle Safety Helmets in Thompson et al. (1989)
y <- matrix(c(17,218,233,758),2,2)
x <- factor(c("yes","no"))
tbn <- glm(y~x,family=binomial)
rsq.n(tbn)
rsq.n(tbn,adj=TRUE)
```

---

rsq.partial — *Partial R-Squared for Generalized Linear Models*

Description

Calculate the coefficient of partial determination, aka partial R^2, for both linear and generalized linear models.

Usage

```
rsq.partial(objF,objR=NULL,adj=FALSE,type=c('v','kl','sse','lr','n'))
```

Arguments

- `objF` — an object of class "lm" or "glm", a result of a call to lm, glm, or glm.nb to fit the full model.
- `objR` — an object of class "lm" or "glm", a result of a call to lm, glm, or glm.nb to fit the reduced model.
- `adj` — logical; if TRUE, calculate the adjusted partial R^2.
- `type` — the type of R-squared:
  'v' (default) – variance-function-based (Zhang, 2016), calling rsq.v;
  'kl' – KL-divergence-based (Cameron and Windmeijer, 1997), calling rsq.kl;
  'sse' – SSE-based (Efron, 1978), calling rsq.sse;
  'lr' – likelihood-ratio-based (Maddala, 1983; Cox and Snell, 1989; Magee, 1990), calling rsq.lr;
  'n' – corrected version of 'lr' (Nagelkerke, 1991), calling rsq.n.

Details

When the fitting object of the reduced model is not specified, the partial R^2 of each term in the model will be calculated.

Value

Returned values include `adjustment` and `partial.rsq`. When `objR` is not `NULL`, `variable.full` and `variable.reduced` are returned; otherwise `variable` is returned.

- `adjustment` — logical; if TRUE, calculate the adjusted partial R^2.
- `variable.full` — all covariates in the full model.
- `variable.reduced` — all covariates in the reduced model.
- `variable` — all covariates in the full model.
- `partial.rsq` — partial R^2 or the adjusted partial R^2.

Author(s)

Dabao Zhang, Department of Statistics, Purdue University

References

Cameron, A. C. and Windmeijer, A. G. (1997) An R-squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics*, 77: 329-342.

Cox, D. R. and Snell, E. J. (1989) *The Analysis of Binary Data*, 2nd ed. London: Chapman and Hall.

Efron, B. (1978) Regression and ANOVA with zero-one data: measures of residual variation. *Journal of the American Statistical Association*, 73: 113-121.

Maddala, G. S. (1983) *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University.

Magee, L. (1990) R^2 measures based on Wald and likelihood ratio joint significance tests. *The American Statistician*, 44: 250-253.

Nagelkerke, N. J. D. (1991) A note on a general definition of the coefficient of determination. *Biometrika*, 78: 691-692.

Zhang, D. (2017). A coefficient of determination for generalized linear models. *The American Statistician*, 71(4): 310-316.

See Also

rsq, pcor.

Examples

```
data(hcrabs)
attach(hcrabs)
y <- ifelse(num.satellites>0,1,0)
bnfit <- glm(y~color+spine+width+weight,family=binomial)
rsq.partial(bnfit)

bnfitr <- glm(y~color+weight,family=binomial)
rsq.partial(bnfit,bnfitr)

quasibn <- glm(y~color+spine+width+weight,family=quasibinomial)
rsq.partial(quasibn)

quasibnr <- glm(y~color+weight,family=binomial)
rsq.partial(quasibn,quasibnr)
```

---

rsq.sse *SSE-Based R-Squared*

Description

The sum-of-squared-errors-based R^2 for generalized linear models.

Usage

```
rsq.sse(fitObj,adj=FALSE)
```

Arguments

fitObj	an object of class "lm" or "glm", usually, a result of a call to lm, glm, or glm.nb.
adj	logical; if TRUE, calculate the adjusted R^2.

Details

This version of R^2 was proposed by Efron (1978). It is calculated on the basis of the formula of the classical R^2.

Value

The R^2 or adjusted R^2.

Author(s)

Dabao Zhang, Department of Statistics, Purdue University

References

Efron, B. (1978) Regression and ANOVA with zero-one data: measures of residual variation. *Journal of the American Statistical Association*, 73: 113-121.

See Also

`rsq`,`rsq.partial`,`pcor`.

Examples

```
data(hcrabs)
attach(hcrabs)
y <- ifelse(num.satellites>0,1,0)
bnfit <- glm(y~color+spine+width+weight,family=binomial)
rsq.sse(bnfit)
rsq.sse(bnfit,adj=TRUE)

psfit <- glm(num.satellites~color+spine+width+weight,family=poisson)
rsq.sse(psfit)
rsq.sse(psfit,adj=TRUE)

# Effectiveness of Bycycle Safety Helmets in Thompson et al. (1989)
y <- matrix(c(17,218,233,758),2,2)
x <- factor(c("yes","no"))
tbn <- glm(y~x,family=binomial)
rsq.sse(tbn)
rsq.sse(tbn,adj=TRUE)
```

---

rsq.v — *Variance-Function-Based R-Squared*

Description

Calculate the variance-function-based R-squared for generalized linear (mixed) models.

Usage

```
rsq.v(fitObj,adj=FALSE,DFUN=NULL)
```

Arguments

fitObj
: an object of class "lm", "glm", "lme", or "glmerMod", usually, a result of a call to lm, glm, glm.nb, glmer, or glmer.nb.

adj
: logical; if TRUE, calculate the adjusted R^2.

DFUN
: function which calculates the directed length along the variance function, and should be specified for quasi models (with family=quasi()).

Details

The R^2 relies on the variance function, and is well-defined for quasi models. It reduces to the classical R^2 when the variance function is constant or linear. For quasi models specified with family=quasi(), DFUN must be specified when the variance function is not constant or linear. For (generalized) linear mixed models, there are three types of R^2 calculated on the basis of observed response values, estimates of fixed effects, and variance components, i.e., model-based R_M^2 (proportion of variation explained by the model in total, including both fixed-effects and random-efffects factors), fixed-effects R_F^2 (proportion of variation explained by the fixed-effects factors), and random-effects R_R^2 (proportion of variation explained by the random-effects factors).

Value

The R^2 or adjusted R^2. For (generalized) linear mixed models,

R_M^2
: proportion of variation explained by the model in total, including both fixed-effects and random-efffects factors.

R_F^2
: proportion of variation explained by the fixed-effects factors.

R_R^2
: proportion of variation explained by the random-effects factors.

Author(s)

Dabao Zhang, Department of Statistics, Purdue University

References

Zhang, D. (2017). A coefficient of determination for generalized linear models. *The American Statistician*, 71(4): 310-316.

Zhang, D. (2020). Coefficients of determination for generalized linear mixed models. *Technical Report*, 20-01, Department of Statistics, Purdue University.

See Also

vresidual, rsq, rsq.glmm, rsq.partial, pcor.

Examples

```
data(hcrabs)
attach(hcrabs)
y <- ifelse(num.satellites>0,1,0)
bnfit <- glm(y~color+spine+width+weight,family=binomial)
```

```
rsq.v(bnfit)
rsq.v(bnfit,adj=TRUE)

quasibn <- glm(y~color+spine+width+weight,family=quasibinomial)
rsq.v(quasibn)
rsq.v(quasibn,adj=TRUE)

# Generalized linear mixed models
require(lme4)
data(cbpp)
glmm1 <- glmer(cbind(incidence,size-incidence)~period+(1|herd),data=cbpp,family=binomial)
rsq.v(glmm1)
```

---

simglm *Simulate Data from Generalized Linear Models*

Description

Simulate data from linear and generalized linear models. Only the first covariate truely affects the response variable with coefficient equal to `lambda`.

Usage

```
simglm(family=c("binomial", "gaussian", "poisson","Gamma"),lambda=3,n=50,p=3)
```

Arguments

`family`	the family of the distribution.
`lambda`	size of the coefficient of the first covariate.
`n`	the sample size.
`p`	the number of covarites.

Details

The first covariate takes 1 in half of the observations, and 0 or -1 in the other half. When `lambda` gets larger, it is supposed to easier to predict the response variable.

Value

Returned values include `yx` and `beta`.

`yx`	a data frame including the response `y` and covariates `x.1`, `x.2`, and so on.
`beta`	true values of the regression coefficients.

Author(s)

Dabao Zhang, Department of Statistics, Purdue University

References

Zhang, D. (2017). A coefficient of determination for generalized linear models. *The American Statistician*, 71(4): 310-316.

See Also

rsq, rsq.partial, pcor.

Examples

```
# Poisson Models
sdata <- simglm(family="poisson",lambda=4)
fitf <- glm(y~x.1+x.2+x.3,family=poisson,data=sdata$yx)
rsq(fitf)  # type='v'

fitr <- glm(y~x.2+x.3,family=poisson,data=sdata$yx)
rsq(fitr)  # type='v'
rsq(fitr,type='kl')
rsq(fitr,type='lr')
rsq(fitr,type='n')

pcor(fitr)  # type='v'
pcor(fitr,type='kl')
pcor(fitr,type='lr')
pcor(fitr,type='n')

# Gamma models with shape=100
n <- 50
sdata <- simglm(family="Gamma",lambda=4,n=n)
fitf <- glm(y~x.1+x.2+x.3,family=Gamma,data=sdata$yx)
rsq(fitf)  # type='v'
rsq.partial(fitf)  # type='v'

fitr <- glm(y~x.2,family=Gamma,data=sdata$yx)
rsq(fitr)  # type='v'
rsq(fitr,type='kl')
rsq(fitr,type='lr')
rsq(fitr,type='n')

# Likelihood-ratio-based R-squared
y <- sdata$yx$y
yhatr <- fitr$fitted.values
fit0 <- update(fitr,.~1)
yhat0 <- fit0$fitted.values
llr <- sum(log(dgamma(y,shape=100,scale=yhatr/100)))
ll0 <- sum(log(dgamma(y,shape=100,scale=yhat0/100)))

# Likelihood-ratio-based R-squared
1-exp(-2*(llr-ll0)/n)

# Corrected likelihood-ratio-based R-squared
(1-exp(-2*(llr-ll0)/n))/(1-exp(2*ll0/n))
```

simglmm *Simulate Data from Generalized Linear Mixed Models*

Description

Simulate data from linear and generalized linear mixed models. The coefficients of the two covariate are specified by `beta`.

Usage

```
simglmm(family=c("binomial","gaussian","poisson","Gamma"),beta=c(2,0),tau=1,n=200,m=10)
```

Arguments

`family`	the family of the distribution.
`beta`	regression coefficients (excluding the intercept which is set as zero).
`tau`	the variance of the random intercept.
`n`	the sample size.
`m`	the number of groups.

Details

The first covariate takes 1 in half of the observations, and 0 or -1 in the other half. When `beta` gets larger, it is supposed to easier to predict the response variable.

Value

Returned values include `yx`, `beta`, and `u`.

`yx`	a data frame including the response `y` and covariates `x1`, `x2`, and so on.
`beta`	true values of the regression coefficients.
`u`	the random intercepts.

Author(s)

Dabao Zhang, Department of Statistics, Purdue University

References

Zhang, D. (2020). Coefficients of determination for generalized linear mixed models. *Technical Report*, 20-01, Department of Statistics, Purdue University.

See Also

`rsq`,`rsq.lmm`,`rsq.glmm`.

Examples

```
require(lme4)

# Linear mixed models
gdata <- simglmm(family="gaussian")
lmm1 <- lmer(y~x1+x2+(1|subject),data=gdata$yx)
rsq(lmm1)

# Generalized linear mixed models
bdata <- simglmm(family="binomial",n=400,m=20)
glmm1 <- glmer(y~x1+x2+(1|subject),family="binomial",data=bdata$yx)
rsq(glmm1)
```

`toxo` *Toxoplasmosis Test in El Salvador*

Description

Recorded are the numbers of subjects testing positive for toxoplasmosis in 34 cities of El Salvador.

Usage

```
data("toxo")
```

Format

A data frame with the test results in 34 cities of El Salvador, includingthe following 4 variables.

`city` index of each city.

`positive` the number of subjects testing positive for toxoplasmosis.

`nsubs` the total number of subjects tested.

`rainfall` annual rainfall (mm) in home city of subject.

Details

All subjects are between 11 and 15 year old. The data set was abstracted from a larger data set in Rmington et al. (1970).

Author(s)

Dabao Zhang, Department of Statistics, Purdue University

Source

Efron, B. (1978). Regression and ANOVA with zero-one data: measures of residual variation. *JASA*, 73: 113-121.

References

Remington, J.S., Efron, B., Cavanaugh, E., Simon, H.J., and Trejos, A. (1970). Studies on toxoplasmosis in El Salvador, prevalence and incidence of toxoplasmosis as measured by the Sabin-Feldman Dye test. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 64: 252-267.

See Also

rsq,rsq.partial,pcor,simglm.

Examples

```
data(toxo)
summary(toxo)
attach(toxo)

toxofit<-glm(cbind(positive,nsubs-positive)~rainfall+I(rainfall^2)+I(rainfall^3),family=binomial)

rsq(toxofit)
rsq(toxofit,adj=TRUE)
rsq.partial(toxofit)

detach(toxo)
```

---

vresidual *Variance-Function-Based Residuals*

Description

Calculate the variance-function-based residuals for generalized linear models, which are used to calculate the variance-function-based R-squared.

Usage

```
vresidual(y,yfit,family="binomial",DFUN=NULL,theta=1)
```

Arguments

y	a vector of observed values.
yfit	a vector of fitted values.
family	family of the distribution.
DFUN	function which calculates the directed distance along the variance function (specified by family by default).
theta	extra parameter for ceterain family like negative binomial.

Details

The calcualted residual relies on the variance function, and is well-defined for quasi models. It reduces to the classical residual when the variance function is constant or linear. Note that only one of DFUN and family (when the family is not "quasi") is required to specify. When the family is "quasi" and the variance is not constant, DFUN should be specified.

Value

Variance-function-based residuals.

Author(s)

Dabao Zhang, Department of Statistics, Purdue University

References

Zhang, D. (2017). A coefficient of determination for generalized linear models. *The American Statistician*, 71(4): 310-316.

See Also

rsq.v, rsq.

Examples

```
data(hcrabs)
attach(hcrabs)
y <- ifelse(num.satellites>0,1,0)
bnfit <- glm(y~color+spine+width+weight,family="binomial")
vresidual(y,bnfit$fitted.values,family="binomial")

# Effectiveness of Bycycle Safety Helmets in Thompson et al. (1989)
y <- matrix(c(17,218,233,758),2,2)
x <- factor(c("yes","no"))
tbn <- glm(y~x,family="binomial")
vresidual(y,tbn$fitted.values,family="binomial")
```

Index