Package ‘ICglm’

October 12, 2022

Title Information Criteria for Generalized Linear Regression

Version 0.1.0

Maintainer Fatih Saglam <fatih.saglam@omu.edu.tr>

Description Calculate various information criteria in literature for ``lm'' and ``glm'' objects.

License MIT + file LICENSE

Imports stats, Matrix

Suggests MASS

Encoding UTF-8

LazyData false

RoxygenNote 7.1.1

NeedsCompilation no

Author Fatih Saglam [aut, cre] (<https://orcid.org/0000-0002-2084-2008>),
Emre Dunder [aut]

Repository CRAN

Date/Publication 2021-11-11 19:10:08 UTC

R topics documented:

AIC *Akaike Information Criterion*

Description

Calculates Akaike Information Criterion (AIC) and its variants for "lm" and "glm" objects.

Usage

```
AIC(model)

AIC4(model)
```

Arguments

`model`	a "lm" or "glm" object

Details

AIC (Akaike, 1973) is calculated as

$$-2LL(theta) + 2k$$

and AIC4 (Bozdogan, 1994) as

$$-2LL(theta) + 2klog$$

Value

AIC or AIC4 measurement of the model

References

Akaike H., 1973. Maximum likelihood identification of Gaussian autoregressive moving average models. Biometrika, 60(2), 255-265.

Bozdogan, H. 1994. Mixture-model cluster analysis using model selection criteria and a new informational measure of complexity. In Proceedings of the first US/Japan conference on the frontiers of statistical modeling: An informational approach, 69–113. Dordrecht: Springer.

Examples

```
x1 <- rnorm(100, 3, 2)
x2 <- rnorm(100, 5, 3)
x3 <- rnorm(100, 67, 5)
err <- rnorm(100, 0, 4)

## round so we can use it for Poisson regression
```

```
y <- round(3 + 2*x1 - 5*x2 + 8*x3 + err)

m1 <- lm(y~x1 + x2 + x3)
m2 <- glm(y~x1 + x2 + x3, family = "gaussian")
m3 <- glm(y~x1 + x2 + x3, family = "poisson")

AIC(m1)
AIC(m2)
AIC(m3)
AIC4(m1)
AIC4(m2)
AIC4(m3)
```

BIC *Bayesian Information Criterion*

Description

Calculates Bayesian Information Criterion (BIC) and its variants (BICadj, BICQ) for "lm" and "glm" objects.

Usage

```
BIC(model)

BICadj(model)

BICQ(model, q = 0.25)
```

Arguments

`model`	a "lm" or "glm" object
`q`	adjustment parameter for `BICQ`. Default is 0.25.

Details

BIC (Schwarz, 1978) is calculated as

$$-2LL(theta) + klog(n)$$

Adjusted BIC (Dziak et al., 2020) as

$$-2LL(theta) + klog(n/2pi)$$

and BICQ (Xu, 2010) as

$$-2LL(theta) + klog(n) - 2klog(q/(1 - q))$$

.

Value

BIC, BICadj or BICQ measurement of the model

References

Dziak, J. J., Coffman, D. L., Lanza, S. T., Li, R., & Jermiin, L. S. (2020). Sensitivity and specificity of information criteria. Briefings in bioinformatics, 21(2), 553-565.

Xu, C. (2010). Model Selection with Information Criteria.

Schwarz, G. 1978. Estimating the dimension of a model The Annals of Statistics 6 (2), 461–464. <doi:10.1214/aos/1176344136>

Examples

```
x1 <- rnorm(100, 3, 2)
x2 <- rnorm(100, 5, 3)
x3 <- rnorm(100, 67, 5)
err <- rnorm(100, 0, 4)

## round so we can use it for Poisson regression
y <- round(3 + 2*x1 - 5*x2 + 8*x3 + err)

m1 <- lm(y~x1 + x2 + x3)
m2 <- glm(y~x1 + x2 + x3, family = "gaussian")
m3 <- glm(y~x1 + x2 + x3, family = "poisson")

BIC(m1)
BIC(m2)
BIC(m3)
BICadj(m1)
BICadj(m2)
BICadj(m3)
```

CAIC *Consistent Akaike's Information Criterion and Consistent Akaike's Information Criterion with Fisher Information*

Description

Consistent Akaike's Information Criterion (CAIC) and Consistent Akaike's Information Criterion with Fisher Information (CAICF) for "lm" and "glm" objects.

Usage

```
CAIC(model)

CAICF(model)
```

Arguments

`model` a "lm" or "glm" object.

Details

CAIC (Bozdogan, 1987) is calculated as

$$-2LL(theta) + k(log(n) + 1)$$

CAICF (Bozdogan, 1987) as

$$-2LL(theta) + 2k + k(log(n)) + log(|F|)$$

F is the Fisher information matrix.

Value

CAIC or CAICF measurement of the model.

References

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. Psychometrika, 52(3), 345-370.

Examples

```
x1 <- rnorm(100, 3, 2)
x2 <- rnorm(100, 5, 3)
x3 <- rnorm(100, 67, 5)
err <- rnorm(100, 0, 4)

## round so we can use it for Poisson regression
y <- round(3 + 2*x1 - 5*x2 + 8*x3 + err)

m1 <- lm(y~x1 + x2 + x3)
m2 <- glm(y~x1 + x2 + x3, family = "gaussian")
m3 <- glm(y~x1 + x2 + x3, family = "poisson")

CAIC(m1)
CAIC(m2)
CAIC(m3)
CAICF(m1)
CAICF(m2)
CAICF(m3)
```

FIC *Fisher Information Criterion*

Description

Calculates Fisher Information Criterion (FIC) for "lm" and "glm" objects.

Usage

```
FIC(model)
```

Arguments

`model`	a "lm" or "glm" object

Details

FIC (Wei, 1992) is calculated as

$$-2LL(theta) + log(|X^T X|)$$

Value

FIC measurement of the model

References

Wei, C. Z. (1992). On predictive least squares principles. The Annals of Statistics, 20(1), 1-42.

Examples

```
x1 <- rnorm(100, 3, 2)
x2 <- rnorm(100, 5, 3)
x3 <- rnorm(100, 67, 5)
err <- rnorm(100, 0, 4)

## round so we can use it for Poisson regression
y <- round(3 + 2*x1 - 5*x2 + 8*x3 + err)

m1 <- lm(y~x1 + x2 + x3)
m2 <- glm(y~x1 + x2 + x3, family = "gaussian")
m3 <- glm(y~x1 + x2 + x3, family = "poisson")

FIC(m1)
FIC(m2)
FIC(m3)
```

GCV *Generalized Cross-Validation*

Description

Calculates Generalized Cross-Validation (GCV) for "lm" and "glm" objects.

Usage

```
GCV(model)
```

Arguments

`model`	a "lm" or "glm" object

Details

GCV (Koc and Bozdogan, 2015) is calculated as

$$RSS/(n(1-k/n))$$

RSS is the residual sum of squares.

Value

GCV measurement of the model

References

Koc, E. K., & Bozdogan, H. (2015). Model selection in multivariate adaptive regression splines (MARS) using information complexity as the fitness function. Machine Learning, 101(1), 35-58.

Examples

```
x1 <- rnorm(100, 3, 2)
x2 <- rnorm(100, 5, 3)
x3 <- rnorm(100, 67, 5)
err <- rnorm(100, 0, 4)

## round so we can use it for Poisson regression
y <- round(3 + 2*x1 - 5*x2 + 8*x3 + err)

m1 <- lm(y~x1 + x2 + x3)
m2 <- glm(y~x1 + x2 + x3, family = "gaussian")
m3 <- glm(y~x1 + x2 + x3, family = "poisson")

GCV(m1)
GCV(m2)
GCV(m3)
```

HBIC *Haughton Bayesian information criterion*

Description

Calculates Haughton Bayesian information criterion (HBIC) for "lm" and "glm" objects.

Usage

```
HBIC(model)
```

Arguments

`model` a "lm" or "glm" object

Details

HBIC (Bollen et al., 2014) is calculated as

$$-2LL(theta) + klog(n/(2pi))$$

Value

HBIC measurement of the model

References

Bollen, K. A., Harden, J. J., Ray, S., & Zavisca, J. (2014). BIC and alternative Bayesian information criteria in the selection of structural equation models. Structural equation modeling: a multidisciplinary journal, 21(1), 1-19.

Examples

```
x1 <- rnorm(100, 3, 2)
x2 <- rnorm(100, 5, 3)
x3 <- rnorm(100, 67, 5)
err <- rnorm(100, 0, 4)

## round so we can use it for Poisson regression
y <- round(3 + 2*x1 - 5*x2 + 8*x3 + err)

m1 <- lm(y~x1 + x2 + x3)
m2 <- glm(y~x1 + x2 + x3, family = "gaussian")
m3 <- glm(y~x1 + x2 + x3, family = "poisson")

HBIC(m1)
HBIC(m2)
HBIC(m3)
```

HQIC *Hannan-Quinn Information Criterion*

Description

Calculates Hannan-Quinn Information Criterion (HQIC) for "lm" and "glm" objects.

Usage

```
HQIC(model)
```

Arguments

`model`	a "lm" or "glm" object

Details

HQIC (Hannan and Quinn, 1979) is calculated as

$$-2LL(theta) + 2klog(log(n))$$

Value

HQIC measurement of the model

References

Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. Journal of the Royal Statistical Society: Series B (Methodological), 41(2), 190-195.

Examples

```
x1 <- rnorm(100, 3, 2)
x2 <- rnorm(100, 5, 3)
x3 <- rnorm(100, 67, 5)
err <- rnorm(100, 0, 4)

## round so we can use it for Poisson regression
y <- round(3 + 2*x1 - 5*x2 + 8*x3 + err)

m1 <- lm(y~x1 + x2 + x3)
m2 <- glm(y~x1 + x2 + x3, family = "gaussian")
m3 <- glm(y~x1 + x2 + x3, family = "poisson")

HQIC(m1)
HQIC(m2)
HQIC(m3)
```

IBIC *Information Matrix-Based Information Criterion*

Description

Calculates Information Matrix-Based Information Criterion (IBIC) for "lm" and "glm" objects.

Usage

```
IBIC(model)
```

Arguments

`model`	a "lm" or "glm" object

Details

IBIC (Bollen et al., 2012) is calculated as

$$-2LL(theta) + klog(n/(2pi)) + log(|F|)$$

F is the fisher information matrix.

While calculating the Fisher information matrix (F), we used the joint parameters ($beta, sigma^2$) of the models.

Value

IBIC measurement of the model

References

Bollen, K. A., Ray, S., Zavisca, J., & Harden, J. J. (2012). A comparison of Bayes factor approximation methods including two new methods. Sociological Methods & Research, 41(2), 294-324.

Examples

```
x1 <- rnorm(100, 3, 2)
x2 <- rnorm(100, 5, 3)
x3 <- rnorm(100, 67, 5)
err <- rnorm(100, 0, 4)

## round so we can use it for Poisson regression
y <- round(3 + 2*x1 - 5*x2 + 8*x3 + err)

m1 <- lm(y~x1 + x2 + x3)
m2 <- glm(y~x1 + x2 + x3, family = "gaussian")
m3 <- glm(y~x1 + x2 + x3, family = "poisson")
```

```
IBIC(m1)
IBIC(m2)
IBIC(m3)
```

IC *Information Criteria*

Description

Calculates Various Information Criteria for "lm" and "glm" objects.

Usage

```
IC(
  model,
  criteria = c("AIC", "BIC", "CAIC", "KIC", "HQIC", "FIC", "ICOMP_IFIM_C1",
    "ICOMP_PEU_C1", "ICOMP_PEU_LN_C1", "CICOMP_C1"),
  ...
)
```

Arguments

`model` a "lm" or "glm" object or object list

`criteria` a vector of criteria names. Can be set to respective numbers. Possible criteria names at the moment are:
1 = "AIC"
2 = "AIC4"
3 = "BIC"
4 = "BICadj"
5 = "BICQ"
6 = "CAIC"
7 = "CAICF"
8 = "FIC"
9 = "GCV"
10 = "HBIV"
11 = "GQIC"
12 = "IBIC"
13 = "ICOMP_IFIM_CF"
14 = "ICOMP_IFIM_C1"
15 = "ICOMP_IFIM_C1F"
16 = "ICOMP_IFIM_C1R"
17 = "ICOMP_PEU_CF"
18 = "ICOMP_PEU_C1"
19 = "ICOMP_PEU_C1F"
20 = "ICOMP_PEU_C1R"
21 = "ICOMP_PEU_LN_CF"

22 = "ICOMP_PEU_LN_C1"
23 = "ICOMP_PEU_LN_C1F"
24 = "ICOMP_PEU_LN_C1R"
25 = "CICOMP_CF"
26 = "CICOMP_C1"
27 = "CICOMP_C1F"
28 = "CICOMP_C1R"
29 = "JIC"
30 = "KIC"
31 = "KICC"
32 = "SPBIC"

`...` additional parameters. Currently none.

Details

Calculates Various Information Criteria for "lm" and "glm" objects. `model` can be a list. If it is a list, function returns a matrix of selected information criteria for all models.

Value

Information criteria of the model(s) for selected criteria

Examples

```
x1 <- rnorm(100, 3, 2)
x2 <- rnorm(100, 5, 3)
x3 <- rnorm(100, 67, 5)
err <- rnorm(100, 0, 4)

## round so we can use it for Poisson regression
y <- round(3 + 2*x1 - 5*x2 + 8*x3 + err)

m1 <- lm(y~x1 + x2 + x3)
m2 <- glm(y~x1 + x2 + x3, family = "gaussian")
m3 <- glm(y~x1 + x2 + x3, family = "poisson")

IC(model = m1, criteria = 1:32)
IC(model = list(lm = m1,
               glm = m2,
               glm_pois = m3), criteria = 1:32)
```

`ICOMP` *Informational Complexity*

Description

These functions calculates Informational Complexity (ICOMP) variants for "lm" and "glm" objects.

Usage

```
ICOMP(model, type = "IFIM", C = "C1")

ICOMP_IFIM_CF(model)

ICOMP_IFIM_C1(model)

ICOMP_IFIM_C1F(model)

ICOMP_IFIM_C1R(model)

ICOMP_PEU_CF(model)

ICOMP_PEU_C1(model)

ICOMP_PEU_C1F(model)

ICOMP_PEU_C1R(model)

ICOMP_PEU_LN_CF(model)

ICOMP_PEU_LN_C1(model)

ICOMP_PEU_LN_C1F(model)

ICOMP_PEU_LN_C1R(model)

CICOMP_CF(model)

CICOMP_C1(model)

CICOMP_C1F(model)

CICOMP_C1R(model)
```

Arguments

`model`	a "lm" or "glm" object
`type`	type of ICOMP. Available types are "IFIM", "PEU", "PEU_LN" and "CICOMP". Default is "IFIM".
`C`	type of complexity. Available types are "CF", "C1", "C1F" and "C1R". Default is "C1".

Details

ICOMP(IFIM) (Bozdogan, 2003) is calculated as

$$-2LL(theta) + 2C(F^{-1})$$

ICOMP(IFIM-peu) (Koc and Bozdogan, 2015) as

$$-2LL(theta) + k + 2C(F^{-1})$$

ICOMP(IFIM-peuln) (Bozdogan, 2010) as

$$-2LL(theta) + k + 2log(n)C(F^{-1})$$

and CICOMP (Pamukcu et al., 2015) as

$$-2LL(theta) + k(log(n) + 1) + 2C(F^{-1})$$

F is the fisher information matrix. F^{-1} is the reverse Fisher information matrix. C is the complexity measure. Four variants are available:

C_1 (Bozdogan, 2010) is

$$C_1(F^{-1}) = s/2 * log(lambda_a/lambda_g)$$

C_F (Bozdogan, 2010) is

$$C_F(F^{-1}) = 1/s * sum_i^s(lambda_i - lambda_a)$$

C_1F (Bozdogan, 2010) is

$$C_1F(F^{-1}) = 1/(4lambda_a^2) * sum_i^s(lambda_i - lambda_a)$$

C_1R (Bozdogan, 2000) is

$$C_1R(F^{-1}) = 1/2 * log(|R|)$$

Here, R is the correlation matrix of the model, $lambda_1, ..., lambda_s$ are eigenvalues of F, $lambda_a$ and $lambda_g$ are arithmetic and geometric mean of eigenvalues of F, respectively. s is the dimension of F. While calculating the Fisher information matrix (F), we used the joint parameters ($beta, sigma^2$) of the models. In $C1R(.)$ function, we utilized the usual variance-covariance matrix $Cov(beta)$ of the models. beta is the vector of regression coefficients.

Value

Informational Complexity measurement of the model

References

Bozdogan, H. (2003). Intelligent Statistical Data Mining with Information Complexity and Genetic Algorithms Hamparsum Bozdogan University of Tennessee, Knoxville, USA. In Statistical data mining and knowledge discovery (pp. 47-88). Chapman and Hall/CRC.

Koc, E. K., & Bozdogan, H. (2015). Model selection in multivariate adaptive regression splines (MARS) using information complexity as the fitness function. Machine Learning, 101(1), 35-58.

Bozdogan, H. (2010). A new class of information complexity (ICOMP) criteria with an application to customer profiling and segmentation. İstanbul Üniversitesi İşletme Fakültesi Dergisi, 39(2), 370-398.

Pamukçu, E., Bozdogan, H., & Çalık, S. (2015). A novel hybrid dimension reduction technique for undersized high dimensional gene expression data sets using information complexity criterion for cancer classification. Computational and mathematical methods in medicine, 2015.

Bozdogan, H. (2000). Akaike's information criterion and recent developments in information complexity. Journal of mathematical psychology, 44(1), 62-91.

Examples

```
x1 <- rnorm(100, 3, 2)
x2 <- rnorm(100, 5, 3)
x3 <- rnorm(100, 67, 5)
err <- rnorm(100, 0, 4)

## round so we can use it for Poisson regression
y <- round(3 + 2*x1 - 5*x2 + 8*x3 + err)

m1 <- lm(y~x1 + x2 + x3)
m2 <- glm(y~x1 + x2 + x3, family = "gaussian")
m3 <- glm(y~x1 + x2 + x3, family = "poisson")

ICOMP_IFIM_CF(m1)
ICOMP_IFIM_CF(m2)
ICOMP_IFIM_CF(m3)
CICOMP_C1(m1)
CICOMP_C1(m2)
CICOMP_C1(m3)
ICOMP(m1, type = "PEU", C = "C1R")
```

`JIC` *Joint Information Criterion*

Description

Joint Information Criterion (JIC) for "lm" and "glm" objects.

Usage

```
JIC(model)
```

Arguments

`model` a "lm" or "glm" object

Details

JIC (Rahman and King, 1999) is calculated as

$$-2LL(theta) + 1/2 * (klog(n) - nlog(1 - k/n))$$

Value

JIC measurement of the model

References

Rahman, M. S., & King, M. L. (1999). Improved model selection criterion. Communications in Statistics-Simulation and Computation, 28(1), 51-71.

Examples

```
x1 <- rnorm(100, 3, 2)
x2 <- rnorm(100, 5, 3)
x3 <- rnorm(100, 67, 5)
err <- rnorm(100, 0, 4)

## round so we can use it for Poisson regression
y <- round(3 + 2*x1 - 5*x2 + 8*x3 + err)

m1 <- lm(y~x1 + x2 + x3)
m2 <- glm(y~x1 + x2 + x3, family = "gaussian")
m3 <- glm(y~x1 + x2 + x3, family = "poisson")

JIC(m1)
JIC(m2)
JIC(m3)
```

KIC — *Kullback–Leibler Information Criterion*

Description

Calculates Kullback–Leibler Information Criterion (KIC) and its corrected form (KICC) for "lm" and "glm" objects.

Usage

```
KIC(model)

KICC(model)
```

Arguments

`model`	a "lm" or "glm" object

Details

KIC (Seghouane, 2006) is calculated as

$$-2LL(theta) + 3k$$

and KICC (Seghouane, 2006) is calculated as

$$-2LL(theta) + ((k + 1)(3n - k - 2)) + (k/(n - k))$$

Value

KIC measurement of the model

References

Seghouane, A. K. (2006). A note on overfitting properties of KIC and KICC. Signal Processing, 86(10), 3055-3060.

Examples

```
x1 <- rnorm(100, 3, 2)
x2 <- rnorm(100, 5, 3)
x3 <- rnorm(100, 67, 5)
err <- rnorm(100, 0, 4)

## round so we can use it for Poisson regression
y <- round(3 + 2*x1 - 5*x2 + 8*x3 + err)

m1 <- lm(y~x1 + x2 + x3)
m2 <- glm(y~x1 + x2 + x3, family = "gaussian")
m3 <- glm(y~x1 + x2 + x3, family = "poisson")

KIC(m1)
KIC(m2)
KIC(m3)
KICC(m1)
KICC(m2)
KICC(m3)
```

SPBIC *Scaled Unit Information Prior Bayesian Information Criterion*

Description

Calculates Scaled Unit Information Prior Bayesian Information Criterion (SPBIC) for "lm" and "glm" objects.

Usage

```
SPBIC(model)
```

Arguments

`model`	a "lm" or "glm" object

Details

SPBIC (Bollen et al., 2012) is calculated as

$$-2LL(theta) + k(1 - log(k/(beta^T(Sigma)^{-1}beta)))$$

beta and Sigma are vector and covariance matrix of regression coefficients.

Value

SPBIC measurement of the model

References

Bollen, K. A., Ray, S., Zavisca, J., & Harden, J. J. (2012). A comparison of Bayes factor approximation methods including two new methods. Sociological Methods & Research, 41(2), 294-324.

Examples

```
x1 <- rnorm(100, 3, 2)
x2 <- rnorm(100, 5, 3)
x3 <- rnorm(100, 67, 5)
err <- rnorm(100, 0, 4)

## round so we can use it for Poisson regression
y <- round(3 + 2*x1 - 5*x2 + 8*x3 + err)

m1 <- lm(y~x1 + x2 + x3)
m2 <- glm(y~x1 + x2 + x3, family = "gaussian")
m3 <- glm(y~x1 + x2 + x3, family = "poisson")

SPBIC(m1)
SPBIC(m2)
```

```
SPBIC(m3)
```

Index