Package ‘ConfZIC’

June 14, 2023

Type Package

Title Confidence Envelopes for Model Selection Criteria Based on Minimum ZIC

Version 1.0.1

Depends R (>= 3.5.0)

Imports cmna, stats, ltsa, MuMIn, mvtnorm, utils, tidytable, psych

Description Narrow down the number of models to look at in model selection using the confidence envelopes based on the minimum ZIC (Generalized Information Criteria) values for regression and time series data. Functions involve the computation of multivariate normal-probabilities with covariance matrices based on minimum ZIC inverting the CDF of the minimum ZIC. It involves both the computation of singular and non-singular probabilities as described in Genz (1992) <[https:doi.org/10.2307/1390838]https:doi.org/10.2307/1390838>.

License GPL-2

Encoding UTF-8

RoxygenNote 7.2.1

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

VignetteBuilder knitr

LazyData true

Maintainer I.M.L. Nadeesha Jayaweera <njayaweera@wpi.edu>

NeedsCompilation no

Author I.M.L. Nadeesha Jayaweera [aut, cre] (<https://orcid.org/0000-0002-5918-8613>), A. Alex Trindade [ctb, aut]

Repository CRAN

Date/Publication 2023-06-14 14:02:09 UTC

R topics documented:

Concrete 2
RankReg 2
RankTS 4
regZIC.test 5
Sunspots 7
tsZIC.test 7

Index **9**

`Concrete` *Concrete Compressive Strength Data Set*

Description

Concrete strength is very important in civil engineering and is a highly nonlinear function of age and ingredients. This dataset contains 1030 instances and there are 8 features relevant to concrete strength.

Usage

```
Concrete
```

Format

A data frame with 1030 rows and 8 covariate variables and 1 response variable

Source

https://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength

Examples

```
data(Concrete)
```

`RankReg` *Rank the regression models based on the confidence envelope for minimum ZIC*

Description

Narrow down the number of models to look at in model selection using the confidence envelope based on the minimum ZIC values for regression data. Here, we compute the ZIC values ("AIC", "BIC", or "AICc") for regression data, confidence envelope for the minimum ZIC values for the given confidence limit, and rank the best models which lie in the confidence envelope.

Usage

```
RankReg(data,alphaval=0.95, model_ZIC="AIC")
```

Arguments

`data`	a matrix of n by $(m+1)$ where m is the number of independent variables. First column should be the dependent variable and the rest of the m columns should be the independent variables of the dataset. Maximum of m should be 10.
`alphaval`	confidence limit of the confidence envelope (Default is 0.95).
`model_ZIC`	type of the information criterion, it can be "AIC", "BIC", or "AICc" (Default is the "AIC").

Details

This program involves the computation of multivariate normal-probabilities with covariance matrices based on minimum ZIC inverting the CDF of the minimum ZIC. It involves both the computation of singular and nonsingular probabilities. The methodology is described in Genz (1992).

Let X_j be the ZIC value for the j^{th} fitted model. Compute the cdf values of the minimum ZIC, $F_{X_{(1)}}(\cdot)$ numerically and then obtain the $100 \cdot (1-\alpha)\%$ confidence envelope:

$$CE(\alpha) = F^{-1}_{X_{(1)}}(1-\alpha)$$

See details:

Jayaweera I.M.L.N, Trindade A.A., “How Certain are You in Your Minimum AIC and BIC Values?", Sankhya A (2023+)

Value

A list containing at least the following components.

`Ranked_Models`	A set of top ranked models which lie in the confidence envelop $CE(\alpha)$ (with variables list and the ranked ZIC values ("AIC", "BIC", or "AICc")) for regression data. 0 represents the coefficient while $1, 2, ..., m$ give the corresponding columns of independent variables $X_1, X_2, ..., X_m$ respectively.
`Confidence_Envelope`	gives the confidence envelope $CE(\alpha)$ for the minimum ZIC.
`Confidence_Limit`	the confidence limit, $1-\alpha$.
`Total_Models`	number of total fitted models.

References

Genz, A. (1992). Numerical computation of multivariate normal probabilities. Journal of computational and graphical statistics, 1(2), 141-149.

Examples

```
library("ConfZIC")
data(Concrete)
x=Concrete
Y=x[,9] #dependent variable
#independent variables
X1=x[,1];X2=x[,2];X3=x[,3];X4=x[,4];
X5=x[,5];X6=x[,6];X7=x[,7];X8=x[,8];
mydata=cbind(Y,X1,X2,X3,X4,X5,X6,X7,X8) #data matrix
RankReg(mydata,0.95,"BIC")
```

`RankTS` *Rank the time series (ARMA) models based on the confidence envelope for minimum ZIC*

Description

Narrow down the number of models to look at in model selection using the confidence envelope based on the minimum ZIC values for time series data. Here, we compute the ZIC values ("AIC", "BIC", or "AICc") for time-series data, confidence envelope for the minimum ZIC values for the given confidence limit, and rank the top models which lie in the confidence envelope.

Usage

```
RankTS(x,max.p,max.q,alphaval=0.95,model_ZIC="AIC")
```

Arguments

`x`	a vector of time series data (should be included with the maximum of 1000 data points).
`max.p`	maximum value for AR coefficient.
`max.q`	maximum value for MA coefficient.
`alphaval`	confidence limit $(1-\alpha)$ (Default is 0.95).
`model_ZIC`	type of the information criterion, it can be "AIC", "BIC", or "AICc" (Default is the "AIC").

Details

This program involves the computation of multivariate normal-probabilities with covariance matrices based on minimum ZIC inverting the CDF of the minimum ZIC. It involves both the computation of singular and non-singular probabilities. The methodology is described in Genz (1992).

Let X_j be the ZIC value for the j^{th} fitted model. Compute the cdf values of the minimum ZIC, $F_{X_{(1)}}(\cdot)$ numerically and then obtain the $100 \cdot (1-\alpha)\%$ confidence envelope:

$$CE(\alpha) = F^{-1}_{X_{(1)}}(1-\alpha)$$

See details:

Jayaweera I.M.L.N, Trindade A.A., "How Certain are You in Your Minimum AIC and BIC Values?", Sankhya A (2023+)

Value

a list of ranked models which lies in the confidence envelope, $CE(\alpha)$.

`Ranked_Models` A set of top ranked time series models which lie in the confidence envelope $CE(\alpha)$ (with AR and MA coefficients, ZIC values ("AIC", "BIC", or "AICc")).

`Confidence_Envelope` gives the confidence envelope $CE(\alpha)$ for the minimum ZIC.

`Confidence_Limit` the confidence limit, $1-\alpha$.

`Total_Models` number of total fitted models.

References

Genz, A. (1992). Numerical computation of multivariate normal probabilities. Journal of computational and graphical statistics, 1(2), 141-149.

Examples

```
library("ConfZIC")
data(Sunspots)
x=Sunspots
RankTS(x,max.p=13,max.q=13,0.95,"AICc")
```

`regZIC.test` *Test whether two ZIC values differ significantly based on minimum ZIC for regression data*

Description

Test whether two ZIC values differ significantly based on minimum ZIC for regression data.

Usage

```
regZIC.test(model1,model2,model_ZIC="AIC",data,alpha=0.05)
```

Arguments

- `model1` — an object of class “lm".
- `model2` — an object of class “lm".
- `model_ZIC` — type of the information criterion, it can be "AIC", "BIC", or "AICc" (Default is the "AIC").
- `data` — a matrix of n by $(m+1)$ where m is the number of independent variables.First column should be the dependent variable and the rest of the m columns should be the independent variables of the dataset. Maximum of m should be 10.
- `alpha` — significance level α for the hypothesis testing (Default is 0.05).

Details

Consider the hypothesis: Under the null hypothesis that the two expected discrepancies are equal.

$$H_0 : ZIC_i = ZIC_j, H_1 : ZIC_i \neq ZIC_j$$

$$Z_0 = \frac{(Z\hat{I}C_i - Z\hat{I}C_j) - 0}{\sqrt{SD(ZIC_i, ZIC_j)}} \sim N(0,1)$$

is calculated empirically.

Value

p-value with significance status.

References

Linhart, H. (1988). A test whether two AIC’s differ significantly. South African Statistical Journal, 22(2), 153-161.

Examples

```
library(ConfZIC)
data(Concrete)
x=Concrete
Y=x[,9] #dependent variable
#independent variables
X1=x[,1];X2=x[,2];X3=x[,3];X4=x[,4];
X5=x[,5];X6=x[,6];X7=x[,7];X8=x[,8];
mydata=cbind(Y,X1,X2,X3,X4,X5,X6,X7,X8) #data matrix
model1=lm(Y~X1); model2=lm(Y~X1+X2)
regZIC.test(model1,model2,model_ZIC="BIC",data=mydata,alpha=0.05)
```

`Sunspots` *Number of sunspots, 1770 to 1869*

Description

Number of sunspots, 1770 to 1869

Usage

```
Sunspots
```

Format

Number of sunspots, 1770 to 1869

Source

Brockwell, P. J., & Davis, R. A. (Eds.). (2002). Introduction to time series and forecasting. New York, NY: Springer New York.

Examples

```
data(Sunspots)
```

`tsZIC.test` *Test whether two ZIC values differ significantly based on minimum ZIC for time series data*

Description

Test whether two ZIC values differ significantly based on minimum ZIC for time series data.

Usage

```
tsZIC.test(x,model1,model2,model_ZIC="AIC",alpha=0.05)
```

Arguments

`x`	time series data (maximum of 1000 data points).
`model1`	AR and MA coefficients of Model 1.
`model2`	AR and MA coefficients of Model 2.
`model_ZIC`	type of the information criterion, it can be "AIC", "BIC", or "AICc" (Default is the "AIC").
`alpha`	significance level α for the hypothesis testing (Default is 0.05).

Details

Consider the hypothesis: Under the null hypothesis that the two expected discrepancies are equal.

$$H_0 : ZIC_i = ZIC_j, H_1 : ZIC_i \neq ZIC_j$$

$$Z_0 = \frac{(Z\hat{I}C_i - Z\hat{I}C_j) - 0}{\sqrt{SD(ZIC_i, ZIC_j)}} \sim N(0, 1)$$

is calculated empirically.

Value

p-value with significance status.

References

Linhart, H. (1988). A test whether two AIC's differ significantly. South African Statistical Journal, 22(2), 153-161.

Examples

```
library(ConfZIC)
data(Sunspots)
x=Sunspots
model1=try(arima(x,order=c(1,0,1),method="ML",include.mean=FALSE),silent = TRUE)
model2=try(arima(x,order=c(1,0,0),method="ML",include.mean=FALSE),silent = TRUE)
tsZIC.test(x,model1,model2,model_ZIC="AIC",alpha=0.05)
```

Index

∗ **datasets**
    Concrete, 2
    Sunspots, 7

Concrete, 2

RankReg, 2
RankTS, 4
regZIC.test, 5

Sunspots, 7

tsZIC.test, 7