

Package: AICcPermanova (via r-universe)

October 12, 2024

Title Model Selection of PERMANOVA Models Using AICc

Version 0.0.3

Description Provides tools for model selection and model averaging of PerMANOVA models using Akaike Information Criterion corrected for small sample sizes (AICc) and Information Theoretic criteria principles. The package is built around the PERMANOVA analysis from the 'vegan' package and provides a streamlined workflow for generating and comparing models, obtaining model weights, and summarizing results using model averaging approaches. The methods implemented in this package are based on the practical information- theoretic approach described by Burnham, K. P. and Anderson, D. R. (2002) (<doi:10.1007/b97636>).

License MIT + file LICENSE

Encoding UTF-8

Roxygen list(markdown = TRUE, roclets = c("`namespace", "`rd"))

RoxygenNote 7.2.3

Imports broom, car, data.table, doParallel, dplyr, foreach, furr, future, parallel, stats, stringr, tidy, vegan

URL <https://github.com/Sustainscapes/AICcPerm>,
<https://sustainscapes.github.io/AICcPerm/>

BugReports <https://github.com/Sustainscapes/AICcPerm/issues>

Suggests betapart, covr, knitr, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

VignetteBuilder knitr

Repository <https://sustainscapes.r-universe.dev>

RemoteUrl <https://github.com/sustainscapes/aiccpem>

RemoteRef HEAD

RemoteSha aae87bb44b464783179d31a7f4451a59108ee995

Contents

AICc_permanova2	2
akaike_adjusted_rsq	3
filter_vif	4
fit_models	5
make_models	7
select_models	8
VIF	8

Index	10
--------------	-----------

AICc_permanova2	<i>Calculate AICc for a permutational multivariate analysis of variance (PERMANOVA)</i>
-----------------	---

Description

#' This function calculates the Akaike's Information Criterion (AICc) for a permutational multivariate analysis of variance (PERMANOVA) model. The AICc is a modified version of the Akaike Information Criterion (AIC) that is more appropriate for small sample sizes and high-dimensional models.

Usage

```
AICc_permanova2(adonis2_model)
```

Arguments

adonis2_model An object of class adonis2 from the vegan package

Details

The AICc calculation for a PERMANOVA model is:

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}$$

where AIC is the Akaike Information Criterion, k is the number of parameters in the model (excluding the intercept), and n is the number of observations.

Value

A data frame with the AICc, the number of parameters (k) and the number of observations (N).

References

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). Mixed effects models and extensions in ecology with R. Springer Science & Business Media.

See Also[adonis2](#)**Examples**

```
library(vegan)
data(dune)
data(dune.env)

# Run PERMANOVA using adonis2

Model <- adonis2(dune ~ Management * A1, data = dune.env)

# Calculate AICc
AICc_permanova2(Model)
```

akaike_adjusted_rsq *Akaike-Adjusted R Squared Calculation with Model Averaging*

Description

Calculates the adjusted R squared for each predictor using the Akaike Information Criterion (AIC) and model averaging. AIC is used to compare the performance of candidate models and select the best one. Then, the R squared is adjusted based on the weight of evidence in favor of each model. The final result is a long-format table of variable names and corresponding adjusted R squared values.

Usage

```
akaike_adjusted_rsq(DF)
```

Arguments

DF A data.frame containing the variables to calculate the adjusted R squared for. The data.frame should include the columns: "form", "AICc", "max_vif", "k", "DeltaAICc", "AICWeight", and "N".

Details

The adjusted R squared is calculated as:

$$\text{Adjusted}R^2 = 1 - (RSS/(N - k - 1)) * ((N - 1)/(N - k - 1))$$

where RSS is the residual sum of squares, N is the sample size, and k is the number of predictors. The R squared is adjusted based on the weight of evidence in favor of each model, which is calculated as:

$$w_i = \exp(-0.5 * \text{DeltaAICc}_i) / \text{sum}(\exp(-0.5 * \text{DeltaAICc}))$$

where w_i is the weight of evidence in favor of the i th model, and $\Delta AICc_i$ is the difference in AICc between the i th model and the best model. Model averaging uses the weights to combine the performance of different models in the final calculation of the adjusted R squared.

Value

A data.frame with columns "Variable" and "Full_Akaike_Adjusted_RSq". Each row represents a predictor, and its corresponding adjusted R squared value based on the Akaike-adjusted model averaging process.

Examples

```
library(data.table)
df <- data.table(
  form = c(1, 2, 3),
  AICc = c(10, 20, 30),
  max_vif = c(3, 4, 5),
  k = c(1, 2, 3),
  DeltaAICc = c(2, 5, 8),
  AICWeight = c(0.2, 0.5, 0.3),
  N = c(100, 100, 100),
  A1 = c(0.3, 0.5, NA),
  A2 = c(0.7, NA, 0.2),
  A3 = c(0.2, 0.3, 0.6)
)
akaike_adjusted_rsq(df)
```

filter_vif

Filters out equations with high multicollinearity

Description

This function takes a dataframe with several models and calculates the maximum Variance Inflation Factor (VIF) for a given model. And either filters out the ones with high collinearity or it flags them accordingly

Usage

```
filter_vif(
  all_forms,
  env_data,
  ncores = 2,
  filter = TRUE,
  threshold = 5,
  verbose = TRUE
)
```

Arguments

all_forms	A data frame generated by make_models
env_data	A dataset with the variables described in all_forms
ncores	An integer specifying the number of cores to use for parallel processing
filter	logical, if TRUE it filters out the models with a maximum VIF of high or higher, if FALSE it generates a new column called collinearity, wich will
threshold	A numeric value specifying the threshold for filtering models based on maximum VIF (default is 5)
verbose	logical, defaults TRUE, sends messages about processing times

Value

A data.frame with the models, filtering out the ones with high collinearity or flagging them.

Examples

```
library(vegan)
data(dune)
data(dune.env)
AllModels <- make_models(vars = c("A1", "Moisture", "Manure"))

filter_vif(
  all_forms = AllModels,
  env_data = dune.env
)
```

fit_models

Fit PERMANOVA models and arrange by AICc

Description

This function fits PERMANOVA models for all combinations of variables in a given dataset, and arranges the models by Akaike Information Criterion (AICc) score. The function also calculates the maximum variance inflation factor (max_vif) for each model.

Usage

```
fit_models(
  all_forms,
  com_data,
  env_data,
  method = "bray",
  ncores = 2,
  log = TRUE,
  logfile = "log.txt",
```

```

    multiple = 100,
    strata = NULL,
    verbose = FALSE
  )

```

Arguments

all_forms	A data frame generated by make_models
com_data	A dataset with community presence absence or abundance data, you can also use a dist class file generated from vegan , betapart or other packages
env_data	A dataset with the variables described in all_forms
method	method for distance from vegdist , this will be ignored if com_data is a distance file
ncores	An integer specifying the number of cores to use for parallel processing
log	logical if true, a log file will be generated
logfile	the text file that will be generated as a log
multiple	after how many loops to write a log file
strata	a block variable similar to the use in adonis2
verbose	logical, defaults TRUE, sends messages about processing times

Value

A data.frame with fitted models arranged by AICc, including the formula used, the number of explanatory variables, R2, adjusted R2, and the AICc and max VIF.

References

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Australian Ecology*, 26(1), 32-46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>

Examples

```

## example with dataframe as community data
library(vegan)
data(dune)
data(dune.env)

AllModels <- make_models(vars = c("A1", "Moisture", "Manure"))

fit_models(
  all_forms = AllModels,
  com_data = dune,
  env_data = dune.env
)
## example with distance as community data
library(betapart)
Distance <- beta.pair.abund(dune)
Distance <- Distance$beta.bray.bal

```

```
fit_models(  
  all_forms = AllModels,  
  com_data = Distance,  
  env_data = dune.env  
)
```

make_models*Create models with different combinations of variables*

Description

Generates all possible linear models for a given set of predictor variables using the distance matrix as a response variable. The function allows for the user to specify the maximum number of variables in a model, which can be useful in cases where there are many predictors. The output is a data frame containing all the possible models, which can be passed to the `fit_models` function for fitting using a PERMANOVA approach.

Usage

```
make_models(vars, ncores = 2, k = NULL, verbose = TRUE)
```

Arguments

<code>vars</code>	A character vector of variables to use for modeling
<code>ncores</code>	An integer specifying the number of cores to use for parallel processing
<code>k</code>	maximum number of variables in a model, default is NULL
<code>verbose</code>	logical, defaults TRUE, sends messages about processing times

Value

A data frame containing all the possible linear permanova models

References

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Australian Ecology*, 26(1), 32-46.

Examples

```
make_models(  
  vars = c("A", "B", "C", "D"),  
  ncores = 2, verbose = FALSE  
)  
  
# using k as a way to limit number of variables
```

```
make_models(
  vars = c("A", "B", "C", "D"),
  ncores = 2, k = 2, verbose = FALSE
)
```

select_models	<i>Select models based on AICc and VIF.</i>
---------------	---

Description

This function selects models from a data frame based on the AICc and VIF values. Models with AICc greater than negative infinity and VIF less than or equal to 6 are considered. The difference in AICc values for each model is calculated with respect to the model with the minimum AICc. Models with a difference in AICc less than or equal to the specified delta_aicc value are selected.

Usage

```
select_models(df, delta_aicc = 2)
```

Arguments

df	a data frame containing the models to select from.
delta_aicc	a numeric value specifying the maximum difference in AICc values allowed.

Value

a data frame containing the selected models and the AIC weights.

Examples

```
df <- data.frame(AICc = c(10, 12, 15, 20), max_vif = c(2, 4, 5, 6))
select_models(df)
select_models(df, delta_aicc = 5)
```

VIF	<i>Get Maximum Variance Inflation Factor (VIF) from a Model</i>
-----	---

Description

This function calculates the maximum Variance Inflation Factor (VIF) for a given model. The VIF is a measure of collinearity among predictor variables within a regression model. It quantifies how much the variance of an estimated regression coefficient is increased due to collinearity. A VIF of 1 indicates no collinearity, while values above 1 indicate increasing levels of collinearity. A VIF of 5 or greater is often considered high, indicating a strong presence of collinearity.

Usage

```
VIF(model)
```

Arguments

`model` A regression model, such as those created by `lm`, `glm`, or other similar functions.

Value

The maximum VIF value.

References

- Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley & Sons.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2004). *Applied Linear Statistical Models*. McGraw-Hill/Irwin.
- O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41(5), 673-690.

Examples

```
data("mtcars")  
VIF(lm(mpg ~ ., data = mtcars))
```

Index

* **collinearity**

VIF, [8](#)

* **models**

AICc_permanova2, [2](#)

adonis2, [3](#), [6](#)

AICc_permanova2, [2](#)

akaike_adjusted_rsqa, [3](#)

filter_vif, [4](#)

fit_models, [5](#)

make_models, [5](#), [6](#), [7](#)

select_models, [8](#)

vegdist, [6](#)

VIF, [8](#)