Package ‘mglmn’

July 29, 2020

Type Package
Title Model Averaging for Multivariate GLM with Null Models
Version 0.1.0
Date 2020-7-29
Author Masatoshi Katabuchi and Akihiro Nakamura
Maintainer Masatoshi Katabuchi <mattocci27@gmail.com>
Description Tools for univariate and multivariate generalized linear models with model averaging and null model technique.
License MIT + file LICENSE
URL https://github.com/mattocci27/mglmn
BugReports https://github.com/mattocci27/mglmn/issues
Depends R (>= 3.5)
Imports mvabund, snowfall
Repository CRAN
RoxygenNote 7.1.1
Suggests testthat
NeedsCompilation no
Date/Publication 2020-07-29 10:20:02 UTC

R topics documented:

best.vars ......................................................... 2
capcay .......................................................... 2
maglm ............................................................ 4
make.formula .................................................... 5
mamglm .......................................................... 6
ses.maglm ....................................................... 7
ses.mamglm ..................................................... 9

Index 11
best.vars  

Best variables

Description

Returns variables for the best model based on AIC

Usage

best.vars(x)

Arguments

x  
A list of results of ‘maglm’ and ‘mamglm’

Value

A vector of terms of the best model.

See Also

maglm, mamglm

Examples

# load species composition and environmental data
data(capcay)
adj.sr <- capcay$adj.sr
env_sp <- capcay$env_sp

# to fit a poisson regression model:
res <- maglm(data = env_sp, y = "adj.sr", family = "gaussian")
best.vars(res)

capcay  

Capcay data

Description

Species composition and environmental data from Capricornia Cays

Usage

data(capcay)
Format

A list containing the elements

**abund**  A data frame with 14 observations of abundance of 13 ant species

**adj.sr**  A vector of adjusted species richness of ants based on sample-based rarefaction curves to standardise sampling intensity across sites (see Nakamura et al. 2015 for more details).

**env_sp**  A data frame of 10 environmental variables, which best explained the variation in the matrix of similarity values.

**env_assem**  A data frame of 10 environmental variables, which best explained the variation in the matrix of similarity values.

The data frame **abund** has the following variables:

- **Camponotus.mackayensis**  (numeric) relative abundance of *Camponotus mackayensis*
- **Cardiocondyla.nuda**  (numeric) relative abundance of *Cardiocondyla nuda*
- **Hypoponera.sp..A**  (numeric) relative abundance of *Hypoponera spA*
- **Hypoponera.sp..B**  (numeric) relative abundance of *Hypoponera spB*
- **Iridomyrmex.sp..A**  (numeric) relative abundance of *Iridomyrmex spA*
- **Monomorium.leave**  (numeric) relative abundance of *Monomorium leave*
- **Ochetellus.sp..A**  (numeric) relative abundance of *Ochetellus spA*
- **Paratrechina.longicornis**  (numeric) relative abundance of *Paratrechina longicornis*
- **Paratrechina.sp..A**  (numeric) relative abundance of *Paratrechina spA*
- **Tapinoma.sp..A**  (numeric) relative abundance of *Tapinoma spA*
- **Tetramorium.bicarinatum**  (numeric) relative abundance of *Tetramorium bicarinatum*

The data frame **env_sp** has the following variables:

- **NativePlSp**  (numeric) native plant species richness
- **P.megaAbund**  (numeric) log-transformed relative abundance of *Pheidole megacephala*
- **P.megaPA**  (numeric) presence/absence of *Pheidole megacephala*
- **HumanVisit**  (numeric) presence/absence of frequent human visitation
- **MaxTemp**  (numeric) mean daily maximum temp(degree celsius)
- **Rain4wk**  (numeric) total rainfall in the past 4 weeks (mm)
- **DistContinent**  (numeric) distance to the nearest continent (km)
- **DistNrIs**  (numeric) log-transformed distance to the nearest island (km)
- **Y**  (numeric) Y coordinate
- **XY**  (numeric) X coordinate * Y coordinate

The data frame **env_assem** has the following variables:

- **IslandSize**  (numeric) log-transformed island size (ha)
- **ExoticPlSp**  (numeric) log-transformed exotic plant species richness
- **NativePlSp**  (numeric) native plant species richness
**P.megaPA** (numeric) presence/absence of *Pheidole megacephala*

**HumanVisit** (numeric) presence/absence of frequent human visitation

**Rainsamp** (numeric) log-transformed total rainfall during sampling (mm)

**DistContinent** (numeric) distance to the nearest continent (km)

**DistNrIs** (numeric) log-transformed distance to the nearest island (km)

**Y** (numeric) Y coordinate

**XY** (numeric) X coordinate * Y coordinate

References


---

**Description**

Model averaging for GLM based on information theory.

**Usage**

`maglm(data, y, family, scale = TRUE, AIC.restricted = FALSE)`

**Arguments**

- `data` Data frame, typically of environmental variables. Rows for sites and columns for environmental variables.
- `y` Vector of independent variables.
- `family` the 'family' object used.
- `scale` Whether to scale independent variables (default = TRUE)
- `AIC.restricted` Whether to use AICc (TRUE) or AIC (FALSE) (default = TRUE).

**Value**

A list of results

- `res.table` data frame with "AIC", AIC of the model, "log.L", log-likelihood of the model, "delta.aic", AIC difference to the best model, "wAIC", weighted AIC to the model, "n.vars", number of variables in the model, and each term.
- `importance` vector of relative importance value of each term, calculated as sum of the weighted AIC over all of the model in which the term appears.
- `family` the 'family' object used.
- `scale` Whether to scale independent variables (default = TRUE)
- `AIC.restricted` Whether to use AICc (TRUE) or AIC (FALSE) (default = TRUE).
References

information-theoretic approach. Springer Verlag, New York.
Nakamura, A., C. J. Burwell, C. L. Lambkin, M. Katabuchi, A. McDougall, R. J. Raven, and V. J.

See Also

mamglm, ses.maglm, ses.mamglm

Examples

#load species composition and environmental data
data(capcay)
adj.sr <- capcay$adj.sr
env_sp <- capcay$env_sp

#to fit a regression model:
maglm(data = env_sp, y = "adj.sr", family = "gaussian", AIC.restricted = TRUE)

make.formula

Utility function

Description

Utility function for data manipulation, which is implemented in maglm and mamglm.

Usage

make.formula(lhs, vars.vec, rand.vec = NULL)

Arguments

lhs : Numeric vector of dependent variables.
vars.vec : Character vector of independent variables.
rand.vec : Character vector of random variables (default = NULL).

Value

an object of class "formula"

See Also

maglm, mamglm
Model averaging for multivariate generalized linear models

Description

Model averaging for multivariate GLM based on information theory.

Usage

mamglm(data, y, family, scale = TRUE, AIC.restricted = FALSE)

Arguments

data Data frame, typically of environmental variables. Rows for sites and columns for environmental variables.
y Name of 'mvabund' object (character)
family the 'family' object used.
scale Whether to scale independent variables (default = TRUE)
AIC.restricted Whether to use AICc (TRUE) or AIC (FALSE) (default = TRUE).

Value

A list of results

- res.table data frame with "AIC", AIC of the model, "log.L", log-likelihood of the model, "delta.aic", AIC difference to the best model, "wAIC", weighted AIC to the model, "n.vars", number of variables in the model, and each term.
- importance vector of relative importance value of each term, calculated as the sum of the weighted AIC over all of the models in which the term appears.
- family the 'family' object used.

References


See Also

maglm, ses.maglm, ses.mamglm
Examples

```r
# load species composition and environmental data
library(mvabund)
data(capcay)
# use a subset of data in this example to reduce run time
env_assem <- capcay$env_assem[, 1:5]
freq.abs <- mvabund(log(capcay$abund + 1))

# to fit a gaussian regression model to frequency data:
mamglm(data = env_assem, y = "freq.abs", family = "gaussian")

# to fit a binomial regression model to presence/absence data
pre.abs0 <- capcay$abund
pre.abs0[pre.abs0 > 0] = 1
pre.abs <- mvabund(pre.abs0)
mamglm(data = env_assem, y = "pre.abs", family = "binomial")
```

---

ses.maglm  

*Standardized effect size of relative importance values for mamglm*

### Description

Standardized effect size of relative importance values for model averaging multivariate GLM.

### Usage

```r
ses.maglm(
  data,  
y,  
family,  
scale = TRUE,  
AIC.restricted = TRUE,  
par = FALSE,  
runs = 999
)
```

### Arguments

- **data**: Data frame, typically of environmental variables. Rows for sites and columns for environmental variables.
- **y**: Vector of independent variables.
- **family**: The 'family' object used.
- **scale**: Whether to scale independent variables (default = TRUE)
- **AIC.restricted**: Whether to use AICc (TRUE) or AIC (FALSE) (default = TRUE).
- **par**: Whether to use parallel computing (default = FALSE).
- **runs**: Number of randomizations.
Details

The currently implemented null model shuffles the set of environmental variables across sites, while maintains species composition. Note that the function would take considerable time to execute.

Value

A data frame of results for each term

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>res.obs</td>
<td>Observed importance of terms</td>
</tr>
<tr>
<td>res.rand.mean</td>
<td>Mean importance of terms in null communities</td>
</tr>
<tr>
<td>res.rand.sd</td>
<td>Standard deviation of importance of terms in null communities</td>
</tr>
<tr>
<td>SES</td>
<td>Standardized effect size of importance of terms (= (res.obs - res.rand.mean) / res.rand.sd)</td>
</tr>
<tr>
<td>res.obs.rank</td>
<td>Rank of observed importance of terms vs. null communities</td>
</tr>
<tr>
<td>runs</td>
<td>Number of randomizations</td>
</tr>
</tbody>
</table>

References


See Also

`maglm`, `mamglm`, `ses.mamglm`

Examples

```r
library(mvabund)
# load species composition and environmental data
data(capcay)
adj.sr <- capcay$adj.sr
# use a subset of data in this example to reduce run time
ev_sp <- capcay$env_sp[, 1:5]

# to execute calculations on a single core:
ses.maglm(data = ev_sp, y = "adj.sr", par = FALSE,
          family = "gaussian", runs = 4)

## Not run:
# to execute parallel calculations:
sfInit(parallel = TRUE, cpus = 4)
sfExportAll()
ses.maglm(data = ev_sp, y = "adj.sr", par = TRUE,
          family = "gaussian", runs = 4)
```
ses.mamglm

## End(Not run)

---

### ses.mamglm

*Standardized effect size of relative importance values for mamglm*

**Description**

Standardized effect size of relative importance values for model averaging GLM.

**Usage**

```r
ses.mamglm(
  data,
  y,
  family,
  scale = TRUE,
  AIC.restricted = TRUE,
  par = FALSE,
  runs = 999
)
```

**Arguments**

- `data` Data frame, typically of environmental variables. Rows for sites and columns for environmental variables.
- `y` Name of `mvabund` object (character)
- `family` the `family` object used.
- `scale` Whether to scale independent variables (default = TRUE)
- `AIC.restricted` Wheter to use AICc (TRUE) or AIC (FALSE) (default = TRUE).
- `par` Wheter to use parallel computing (default = FALSE)
- `runs` Number of randomizations.

**Details**

The currently implemented null model shuffles the set of environmental variables across sites, while maintains species composition. Note that the function would take considerable time to execute.

**Value**

A data frame of results for each term

- `res.obs` Observed importance of terms
- `res.rand.mean` Mean importance of terms in null communities
res.rand.sd  Standard deviation of importance of terms in null communities
SES         Standardized effect size of importance of terms (= (res.obs - res.rand.mean) / res.rand.sd)
res.obs.rank Rank of observed importance of terms vs. null communities
runs         Number of randomizations

References


Examples

```r
library(mvabund)
#load species composition and environmental data
data(capcay)
#use a subset of data in this example to reduce run time
env_assem <- capcay$env_assem[, 1:5]
pre.abs0 <- capcay$abund
pre.abs0[pre.abs0 > 0] = 1
pre.abs <- mvabund(pre.abs0)

#to execute calculations on a single core:
es.mamglm(data = env_assem, y = "pre.abs",
par = FALSE, family = "binomial",
AIC.restricted=FALSE,runs=4)

## Not run:
#to execute parallel calculations:
sfInit(parallel = TRUE, cpus = 4)
sfExportAll()
es.mamglm(data = env_assem, y = "pre.abs",
par = TRUE, family = "binomial",
AIC.restricted = FALSE, runs = 4)

## End(Not run)
```
Index

* datasets
  capcay, 2
  best.vars, 2
  capcay, 2
  maglm, 2, 4, 5, 6, 8
  make.formula, 5
  mamglm, 2, 5, 6, 8
  ses.maglm, 5, 6, 7
  ses.mamglm, 5, 6, 8, 9