Package ‘seqhandbook’

October 14, 2022

Type Package
Title Miscellaneous Tools for Sequence Analysis
Version 0.1.0
Depends R (>= 3.5.0), TraMineR
Suggests knitr, rmarkdown, FactoMineR, GDAtools, RColorBrewer,
   TraMineRextras, WeightedCluster, ade4, cluster, questionr,
   rmdformats, dplyr, purrr, ggplot2
VignetteBuilder knitr
Author Nicolas Robette
Maintainer Nicolas Robette <nicolas.robette@uvsq.fr>
License GPL (>= 2)
Encoding UTF-8
LazyData true
NeedsCompilation no
Repository CRAN
Date/Publication 2020-06-29 15:30:06 UTC

R topics documented:

assoc.domains ................................................................. 2
seq2qha ................................................................. 3
seqgimsa ................................................................. 4
assoc.domains

Description

Computes various measures of association between dimensions of multidimensional sequence data.

Usage

assoc.domains(dlist, names, djsa)

Arguments

dlist     A list of dissimilarity matrices or dist objects (see dist), with one element per dimension of the multidimensional sequence data

names     A character vector of the names of the dimensions of the multidimensional sequence data

djsa      A dissimilarity matrix or a dist object (see dist), corresponding to the distances between the multidimensional sequences

Author(s)

Nicolas Robette

References

Examples

```r
library(TraMineR)
data(biofam)

## Building one channel per type of event (left, children or married)
bf <- as.matrix(biofam[, 10:25])
children <- bf==4 | bf==5 | bf==6
married <- bf == 2 | bf== 3 | bf==6
left <- bf==1 | bf==3 | bf==5 | bf==6

## Building sequence objects
child.seq <- seqdef(children)
marr.seq <- seqdef(married)
left.seq <- seqdef(left)

## Using Hamming distance
mcdist <- seqdistmc(channels=list(child.seq, marr.seq, left.seq), method="HAM")
child.dist <- seqdist(child.seq, method="HAM")
marr.dist <- seqdist(marr.seq, method="HAM")
left.dist <- seqdist(left.seq, method="HAM")

## Association between domains
asso <- assoc.domains(list(child.dist, marr.dist, left.dist), c('child','marr','left'), mcdist)
asso
```

---

**seq2qha**

*Recoding sequences for qualitative harmonic analysis*

**Description**

Recodes sequence data into the shape used for qualitative harmonic analysis.

**Usage**

```r
seq2qha(seqdata, periods)
```

**Arguments**

- `seqdata` a sequence object (see `seqdef` function).
- `periods` numeric vector of the first positions of the periods used for recoding

**Value**

A data frame with one column by combination of period and state (i.e. number of columns = number of periods * number of states in the alphabet).
Author(s)
Nicolas Robette

References

Examples
data(trajact)
seqact <- seqdef(trajact)
qha <- seq2qha(seqact, periods=c(1,3,7,12,24))
head(qha)

Description
A data frame describing mothers employment histories from age 14 to 60 and daughters employment histories from the completion of education to 15 years later. Sequences are sampled (N = 400) from "Biographies et entourage" survey (INED, 2001).

Usage
data("seqgimsa")

Format
A data frame with 400 observations and 62 numeric variables. The first 15 variables (prefixed 'f') describe the daughters employment status a given year : 1 = education, 2 = inactivity, 3 = part-time job, 4 = full-time job. The following 47 variables (prefixed 'm') describe the mothers employment status at a given age : 1 = self-employment, 3 = higher level or intermediate occupation, 5 = lower level occupation, 8 = inactivity, 9 = education.

Examples
data(seqgimsa)
str(seqgimsa)
seqilepi

At least one episode in the states

Description
Returns whether each sequence comprises at least one episode in the states.

Usage
seqilepi(seqdata)

Arguments
seqdata a sequence object (see seqdef function).

Author(s)
Nicolas Robette

References

See Also
seqistatd, seqinepi, seqifpos

Examples
```r
data(trajact)
seqact <- seqdef(trajact)
stat <- seqilepi(seqact)
head(stat)
```

seqifpos
First position in each state

Description
Returns the first position in each state.

Usage
seqifpos(seqdata)
Arguments

seqdata a sequence object (see seqdef function).

Author(s)

Nicolas Robette

References


See Also

seqistatd, seqilepi, seqinepi

Examples

```r
data(trajact)
seqact <- seqdef(trajact)
stat <- seqifpos(seqact)
head(stat)
```

<table>
<thead>
<tr>
<th>seqinepi</th>
<th>Number of episodes in each state</th>
</tr>
</thead>
</table>

Description

Returns the number of episodes in the states.

Usage

```r
seqinepi(seqdata)
```

Arguments

- seqdata a sequence object (see seqdef function).

Author(s)

Nicolas Robette

References

seqmds.stress

See Also

seqistatd, seqilepi, seqifpos

Examples

```r
data(trajact)
seqact <- seqdef(trajact)
stat <- seqinepi(seqact)
head(stat)
```

seqmds.stress  Stress measure of multidimensional scaling factors

Description

Computes stress measure of multidimensional scaling data for different number of dimensions of the represented space

Usage

```r
seqmds.stress(seqdist, mds)
```

Arguments

- `seqdist` a dissimilarity matrix or a dist object (see `dist`)
- `mds` a matrix with coordinates in the represented space (dimension 1 in column 1, dimension 2 in column 2, etc.)

Value

A numerical vector of stress values.

Author(s)

Nicolas Robette

References


Examples

```r
data(trajact)
seqact <- seqdef(trajact)
dissim <- seqdist(seqact, method="HAM")
mds <- cmdscale(dissim, k=20, eig=TRUE)
stress <- seqmds.stress(dissim, mds)
plot(stress, type='l', xlab='number of dimensions', ylab='stress')
```
**seqmsa**  
*Sample of marital, parental and residential sequences*

**Description**
A data frame describing respectively the matrimonial, parental and residential status from age 14 to age 35. It’s sampled (N=500) from "Biographies et entourage" survey (INED, 2001).

**Usage**
data("seqmsa")

**Format**
A data frame with 500 observations and 66 variables. The first 22 variables (prefixed 'log') describe the residential status at a given age: 0 = not independent, 1 = independent. The next 22 variables (prefixed 'mat') describe the matrimonial status at a given age: 1 = never been in a relationship, 2 = cohabiting union, 3 = married, 4 = separated. The last 22 variables (prefixed 'nenf') describe the parental status at a given age: 0 = no child, 1 = one child, 2 = two children, 3 = three children or more.

**Examples**
data(seqmsa)  
str(seqmsa)

**seqsmooth**  
*Smoothing sequence data*

**Description**
Smoothing of sequence data, using for each sequence the medoid of the sequences in its neighborhood. The results can be used to get a smoothed index plot.

**Usage**
seqsmooth(seqdata, diss, k=20, r=NULL)

**Arguments**
- **seqdata**: a sequence object (see seqdef function).
- **diss**: a dissimilarity matrix, giving the pairwise distances between sequences.
- **k**: size of the neighborhood. Default is 20.
- **r**: radius of the neighborhood. If NULL (default), the radius is not used for smoothing.
seq_heatmap

Value

A list with the following elements:

- seqdata: a sequence object (see seqdef function)
- R2: pseudo-R2 measure of the goodness of fit of the smoothing
- S2: stress measure of the goodness of fit of the smoothing

Author(s)

Nicolas Robette

References


Examples

data(trajact)
seqact <- seqdef(trajact)
dissim <- seqdist(seqact, method="LCS")
mds <- cmdscale(dissim, k=1)
smoothed <- seqsmooth(seqact, dissim, k=30)$seqdata
seqIplot(smoothed, sortv=mds, xtlab=14:50, with.legend=FALSE, yaxis=FALSE, ylab=NA)

Description

Index plot of state sequences. Sequences are ordered according to the specified dendrogram. The dendrogram is also plotted on the side of the index plot.

Usage

seq_heatmap(seq, tree, with.missing = FALSE, ...)

Arguments

- seq: a state sequence object created with the seqdef function
- tree: a dendrogram of the sequences (an object of class hclust, dendrogram or agnes)
- with.missing: is there a 'missing value' state in the sequences?
- ...: additional parameters sent to heatmap
**Description**

A data frame with sociodemographic variables for a sample of 500 interviewees from "Biographies et entourage" survey (INED, 2001).

**Usage**

```r
data("socdem")
```

**Format**

A data frame with 500 observations on the following 9 variables.

- **annais**: year of birth (numeric)
- **nbenf**: number of children (factor)
- **nbunion**: number of relationships (factor)
- **mereactive**: whether mother was active or not (factor)
- **sexe**: gender (factor)
- **PCS**: occupational category (factor)
- **PCSpere**: occupational category of the father (factor)
- **diplome**: degree (factor)
- **nationalite**: nationality (factor)

**Examples**

```r
data(socdem)
str(socdem)
```
Description

Computes symmetric (or canonical) PLS for two groups of continuous variables

Usage

symPLS(a, b)

Arguments

a | data frame of the first group of continuous variables
b | data frame of the second group of continuous variables

Author(s)

Nicolas Robette, Xavier Bry

References


Description

A data frame describing the employment status from age 14 to age 50. It’s a sample of 500 interviewees from "Biographies et entourage" survey (INED, 2001).

Usage

data("trajact")

Format

A data frame with 500 observations and 37 variables. Each variable is numeric and describes the employment status at a given age: 1 = education, 2 = full-time job, 3 = part-time job, 4 = small jobs, 5 = inactivity, 6 = military service.

Examples

data(trajact)
str(trajact)
Index

* Longitudinal characteristics
  assoc.domains, 2
  seq2qha, 3
  seq1lepi, 5
  seqifpos, 5
  seqinepi, 6
  seqmds.stress, 7
  seqsmooth, 8
  symPLS, 11

* State sequences
  assoc.domains, 2
  seq2qha, 3
  seq1lepi, 5
  seqifpos, 5
  seqinepi, 6
  seqmds.stress, 7
  seqsmooth, 8
  symPLS, 11

* datasets
  seqgimsa, 4
  seqmsa, 8
  socdem, 10
  trajact, 11

agnes, 9
assoc.domains, 2
dendrogram, 9
dist, 2, 7
hclust, 9
heatmap, 9
seq2qha, 3
seq_heatmap, 9
seqdef, 3, 5, 6, 8, 9
seqgimsa, 4
seq1lepi, 5, 6, 7
seqifpos, 5, 6, 7
seqinepi, 5, 6, 6
seqIplot, 10
seqistatd, 5–7
seqmds.stress, 7
seqmsa, 8
seqsmooth, 8
socdem, 10
symPLS, 11
trajact, 11