Package 'LOST'

April 14, 2020

Version 2.0.2

Date 2020-4-13

Title Missing Morphometric Data Simulation and Estimation

Author J. Arbour,
C. Brown

Maintainer J. Arbour <jessica.arbour@mtsu.edu>

Imports gdata, shapes, e1071, pcaMethods, MASS, miscTools, stats, rgl,
geomorph

Description Functions for simulating missing morphometric
data randomly, with taxonomic bias and with anatomical bias. LOST also
includes functions for estimating linear and geometric morphometric data.

License GPL (>= 2)

Depends R (>= 3.5.0)

NeedsCompilation no

Repository CRAN

Date/Publication 2020-04-14 07:00:11 UTC

R topics documented:

LOST-package   *Missing morphometric data simulation and estimation*

Description

LOST includes functions for simulating missing morphometric data randomly, with taxonomic bias and with anatomical bias as described by Brown et al. 2012. This package also includes functions for estimating missing morphometric data based on regression analysis and a function for checking the percentage of missing data in a matrix.

Details

Package:	LOST
Type:	Package
Version:	1.1
Date:	2020-04-13
License:	GPL (>= 2)

Author(s)

J. Arbour and C. Brown

Maintainer: jessica.arbour@mtsu.edu

References

Arbour, J. and Brown, C. 2014. Incomplete specimens in Geometric Morphometric Analyses. *Methods in Ecology and Evolution*

Brown, C., Arbour, J. and Jackson, D. 2012. Testing of the Effect of Missing Data Estimation and Distribution in Morphometric Multivariate Data Analyses. *Systematic Biology* 61(6):941-954.

---

align.missing   *Procrustes superimposition of landmark datasets with some missing values*

Description

This function carries out a generalized procrustes superimposition on all fully complete specimens and produces a consensus configuration (using "Shapes" procGPA). Each incomplete specimen is then individually rotated and aligned with the consensus configuration based on any landmarks are available (using "Shapes" procOPA). Data is returned superimposed.

Usage

```
align.missing(X)
```

Arguments

X   An l X 2 (or 3) X n array of coordinate data, where n is the number of specimens and l is the number of landmarks.

Value

Returns An l X 2 (or 3) X n array of coordinate data

Author(s)

J. Arbour

References

Arbour, J. and Brown, C. 2014. Incomplete specimens in Geometric Morphometric Analyses. *Methods in Ecology and Evolution* 5(1):16-26.

See Also

MissingGeoMorph

Examples

```
data(dacrya)

## make some specimens incomplete
dac.miss<-missing.data(dacrya,remsp=0.2,land.vec=c(1,2,3,4,5,6))

## align all specimens
dac.aligned<-align.missing(dac.miss)
```

---

best.reg	*Estimate missing morphometric data with a highly correlated variable*

Description

Estimates missing morphometric using regression on the most highly correlated morphological variable available

Usage

```
best.reg(x)
```

Arguments

x   A n X m matrix of morphometric data with n specimens and m variables, containing some percentage of missing values input as NA

Value

Returns a n X m matrix containing both the original morphometric values as well as estimates for all previously missing values.

Author(s)

J. Arbour and C. Brown

References

Brown, C., Arbour, J. and Jackson, D. 2012. Testing of the Effect of Missing Data Estimation and Distribution in Morphometric Multivariate Data Analyses. *Systematic Biology* 61(6):941-954.

See Also

[est.reg](#)

---

byclade	*Simulate missing morphometric data with taxonomic bias*

Description

This function simulates higher frequency of missing data points in groups that are less numerically well represented in the whole sample, relative to other group. These groups may represent taxa (as used in Brown et al., 2012), but may also represent any other group of interest (e.g. populations, trials, subsamples, etc.). From a morphometric dataset, this function first selects a number of specimens to have data points removed from at random. A vector containing the number of measurements to remove from each specimen is sorted into descending order. Specimens are then sampled without replacement with a probability relative to the sum of the entire sample sizes divided by the number of specimens its respective group. The order the specimens are sampled determines the number of data points to be removed (i.e. the first to be sampled has the most removed). A complete mathematical description may be found in Brown et al. (2012).

Usage

```
byclade(x, remperc , groups)
```

Arguments

x	A n X m matrix of morphometric data with n specimens and m variables. Or an l X 2 or 3 X n array of geometric morphometric coordinates (2D or 3D), where l is the number of landmarks.
remperc	The percentage of data to be removed from the matrix, expressed as a decimal (ex: 30 percent would be entered as 0.3)
groups	A vector of length n specifying taxonomic group membership as integers (ex: c(1,1,2,2,3,3,...))

Value

returns a matrix or array (depending on input) of morphometric data with missing variables input as 'NA'

Author(s)

J. Arbour and C. Brown

References

Brown, C., Arbour, J. and Jackson, D. 2012. Testing of the Effect of Missing Data Estimation and Distribution in Morphometric Multivariate Data Analyses. *Systematic Biology* 61(6):941-954.

See Also

[missing.data](),[obliterator]()

---

complete.specimens *Remove incomplete specimens from a landmark dataset*

Description

This function takes a dataset containing both complete and incomplete specimens and removes all incomplete specimens.

Usage

```
complete.specimens(dataset, nlandmarks)
```

Arguments

dataset   A n*l X 2 matrix of coordinate data, where n is the number of specimens and l is the number of landmarks. All landmarks from one specimen should be grouped together.

nlandmarks   The number of landmarks per specimen

Value

Returns an c * l X 2 matrix of landmark data, where c is the number of complete specimens and l is the number of landmarks.

Author(s)

J. Arbour

References

Arbour, J. and Brown, C. *In Press.* Incomplete specimens in Geometric Morphometric Analyses. *Methods in Ecology and Evolution*

See Also

[align.missing](#),code[MissingGeoMorph](#)

---

crocs *Crocodile morphometrics*

Description

A linear morphometric dataset featuring 23 cranial measurements from 223 specimens representing 21 crocodilian species.

Usage

    data(crocs)

Format

A n X m dataframe, where n is the number of specimens and m is the number of variables.

Source

http://datadryad.org/resource/doi:10.5061/dryad.m01st7p0

References

Brown, C., Arbour, J. and Jackson, D. 2012. Testing of the Effect of Missing Data Estimation and Distribution in Morphometric Multivariate Data Analyses. *Systematic Biology* 61(6):941-954.

See Also

[obliterator](), [byclade](),[missing.data](),[crocs.landmarks]()

---

crocs.landmarks *Coordinate data for a crocodilian reference skull*

Description

Landmark data for the measurements points on a reference crocodilian skull, for use with the [obliterator]() function

Usage

    data(crocs.landmarks)

Format

A 6 X m dataframe in which each column gives the start and end points for each cranial measurement in the crocs dataset, from a single reference specimen. 3D Coordinates are listed as x1, x2, y1, y2, z1, z2 in each column.

Source

Brown, C., Arbour, J. and Jackson, D. 2012. Testing of the Effect of Missing Data Estimation and Distribution in Morphometric Multivariate Data Analyses. *Systematic Biology* 61(6):941-954.

See Also

obliterator, byclade, missing.data, crocs

---

dacrya	*Landmark data from* Guianacara dacrya

Description

Sixteen landmarks taken from the lateral profile of 73 specimens from the Essequibo and rio Branco drainages, used in the description of *Guianacara dacrya*

Usage

```
data(dacrya)
```

Format

A 16 X 2 X 73 array of geometric morphometric coordinates

Source

Arbour, J. and Lopez-Fernandez, H. 2011. *Guiancara dacrya*, a new species from the rio Branco and Essequibo River drainages of the Guiana Shield (Perciformes: Cichlidae). *Neotropical Ichthyology* 9:87-96.

See Also

align.missing, MissingGeoMorph

---

est.reg	*A-priori size regression for missing data estimation*

Description

Estimates missing data using regression on a designated size variable. Any values of the size variable missing are estimated with the variable best correlated with size.

Usage

```
est.reg(x, col_indep)
```

Arguments

x	A n X m matrix of morphometric data with n specimens and m variables, containing some percentage of missing values input as NA
col_indep	The number of the column in which the independant size variable is stored. This column will be used to estimate missing values in the other columns.

Value

Returns a n X m matrix containing both the original morphometric values as well as estimates for all previously missing values.

Author(s)

J. Arbour and C. Brown

References

Brown, C., Arbour, J. and Jackson, D. 2012. Testing of the Effect of Missing Data Estimation and Distribution in Morphometric Multivariate Data Analyses. *Systematic Biology* 61(6):941-954.

See Also

best.reg

---

flipped   *Reflected Relabelling*

Description

This function carries out reflected relabelling to estimate missing geometric morphometric landmarks using bilateral symmetry following Gunz et al 2009.

A set of 3D landmarks are mirrored and aligned with the original data (using procOPA from package "shapes"). Missing landmarks are interpolated from the mirrored specimen.

Usage

```
flipped(specimen, land.pairs, show.plot = FALSE, axis = 1)
```

Arguments

specimen	An l X 3 matrix of coordinate data, where l is the number of landmarks. Some data should be missing and designated with NA.
land.pairs	A 2 column matrix, each row should contain row numbers (from matrix specimen) indicating bilateral pairs of landmarks. Unpaired landmarks do not need to be included. See also bilateral symmetry analyses in package "geomorph".
show.plot	Optionally plot the specimen using plot3d from rgl. Estimated landmarks are given in red. Defaults to FALSE.
axis	Which axis should be mirrored across. Default is x (1).

Value

Returns a l X 3 matrix of landmarks.

Author(s)

J. Arbour

References

Gunz P., Mitteroecker P., Neubauer S., Weber G., Bookstein F. 2009. Principles for the virtual reconstruction of hominin crania. Journal of Human Evolution 57:48-62.

See Also

[MissingGeoMorph](#)

---

how.many.missing *Calculate the percentage of missing morphometric data*

Description

Calculates the percentage of morphometric data points that have been replaced with 'NA' by functions such as `missing.data`, `byclade` or `obliterator` from LOST. Used to verify the amount of missing data inputted into complete morphometric matrices.

Usage

```
how.many.missing(x)
```

Arguments

x   A n X m matrix of morphometric data with n specimens and m variables, or a or l X 2(or 3) array of geometric morphometric data containing some percentage of missing data

Value

Returns the percentage (as a decimal) of missing data points present in x

Author(s)

J. Arbour and C. Brown

References

Brown, C., Arbour, J. and Jackson, D. 2012. Testing of the Effect of Missing Data Estimation and Distribution in Morphometric Multivariate Data Analyses. *Systematic Biology* 61(6):941-954.

See Also

missing.data

---

missing.data   *Randomly input missing data points*

Description

Randomly replaces a set percentage of data points in a matrix of morphometric measurements with NA to simulate missing data. This is function RMD from Brown et al (2012). The amount of missing data can be chosen as an overall percentage of data (simple morphometric data) or specimens and can be constrained to a set of landmarks (for landmarks).

Usage

```
missing.data(x, remperc, remsp = NULL, land.vec = NULL, land.identity = NULL)
```

Arguments

x	A n X m matrix of morphometric data with n specimens and m variables. Or an array of geometric morphometrics landmarks (l X m X n)
remperc	The percentage of data to be removed from the matrix or array, expressed as a decimal (ex: 30 percent would be entered as 0.3)
remsp	The percentage of specimens to be removed from the array, expressed as a decimal (ex: 30 percent would be entered as 0.3)
land.vec	The number of landmarks to remove per specimen in an array. This can be a single value or vector with unique or repeating values.
land.identity	A vector to constrain the landmarks to chose from when assigning missing data. The values correspond to row numbers in an array.

Value

Returns a n X m matrix or l X m X n array of morphometric data with missing variables input as NA

Author(s)

J. Arbour and C. Brown

References

Brown, C., Arbour, J. and Jackson, D. 2012. Testing of the Effect of Missing Data Estimation and Distribution in Morphometric Multivariate Data Analyses. *Systematic Biology* 61(6):941-954.

See Also

byclade,obliterator

Examples

```
data(dacrya)

#### remove 1 to 6 landmarks from 20% of specimens
dac.miss<-missing.data(dacrya,remsp=0.2,land.vec=c(1,2,3,4,5,6))
dac.miss
```

---

missing.specimens   *Simulate incomplete specimens*

Description

Randomly selects a pre-determined number of specimens from a landmark dataset (2D or 3D) and removes some of their landmarks.

Usage

```
missing.specimens(dataset, nspremove, nldremove, nlandmarks)
```

Arguments

dataset	A n*l X 2 (or 3) matrix of coordinate data, where n is the number of specimens and l is the number of landmarks. All landmarks from one specimen should be grouped together.
nspremove	The number of specimens which should have landmarks removed.
nldremove	The number of landmarks to remove per specimen. This may be a single value or a vector of values, none of which can be >nlandmarks. If a vector is given, for each specimen selected, the function will randomly select a value from the vector and remove that many landmarks.
nlandmarks	The number of landmarks per specimen

Value

Returns an n * l X 2 (or 3) matrix with some complete and some incomplete specimens.

Author(s)

J. Arbour

References

Arbour, J. and Brown, C. 2014. Incomplete specimens in Geometric Morphometric Analyses. *Methods in Ecology and Evolution* 5(1):16-26.

See Also

[align.missing](), {code[MissingGeoMorph]()

MissingGeoMorph *Estimate missing landmark data*

Description

This function provides several options for estimating landmark data (details of which can be found in the references below). The function first aligns the landmarks using Procrustes superimposition (`align.missing`). Both 2D and 3D coordinates can be accommodated.

Usage

```
MissingGeoMorph(x, method = "BPCA", original.scale = FALSE)
```

Arguments

x   A n* l X 2 matrix (2D data only) or an l X m X n array (2D or 3D data) of coordinate data, where n is the number of specimens and l is the number of landmarks, and m is the number of dimensions. All landmarks from one specimen should be grouped together. Missing values should be given as NA

method   Four methods are provided for estimating missing landmark data: 1) "BPCA" - Bayesian principal component analysis, 2) "mean" - mean substitution, 3) "reg" - values are estimated based on the most strongly correlated variable available, and 4) "TPS" - thin plate spline interpolation (only available for 2D). See Arbour and Brown (2014) for a comparison of the performance of each of these methods.

original.scale Rescale and translate the data back to its original size (TRUE) or leave it in the rescaled, superimposed configuration (FALSE)

Value

Returns an n * l X 2 (or 3) matrix of coordinate data, with missing values imputed. Landmarks have been aligned and are given in the original shape space.

Author(s)

J. Arbour

References

Arbour, J. and Brown, C. 2014. Incomplete specimens in Geometric Morphometric Analyses. *Methods in Ecology and Evolution* 5(1):16-26.

Brown, C., Arbour, J. and Jackson, D. 2012. Testing of the Effect of Missing Data Estimation and Distribution in Morphometric Multivariate Data Analyses. *Systematic Biology* 61(6):941-954.

See Also

`align.missing`, `missing.specimens`

obliterator	*Simulate missing morphometric data with anatomical bias*

Description

This function simulates the effect of proximity between measurements in morphometric data on the distribution of missing values. This attempts to replicate specimens showing regional incompleteness. From a morphometric dataset, this function selects a number of specimens to have data points removed from and a number of measurements to remove from each of these specimens based on a random distribution of missing data. For each specimen, this function randomly selects one starting data point for removal. All subsequent data points have a probability of removal that is proportional to the inverse of the distance to all previously removed data points, based on a reference set of landmarks (matrix 'distances'). For a complete mathematical description see Brown et al. (2012). See function obliteratorGM for the geometric morphometric implementation.

Usage

```
obliterator(x, remperc, landmarks, expo=1)
```

Arguments

x	A n X m matrix of morphometric data with n specimens and m variables
remperc	The percentage of data to be removed from the matrix, expressed as a decimal (ex: 30 percent would be entered as 0.3)
landmarks	A 6 X m matrix that includes the start and end points (landmarks) for each morphometric measurement from a reference specimen (3D). The data in each column is ordered as x1,x2,y1,y2,z1,z2. See example crocs.landmarks
expo	An optional term for raising the denominator to an exponent, to increase or decrease the severity of the anatomical bias

Value

Returns a n X m matrix of morphometric data with missing variables input as NA

Author(s)

J. Arbour and C. Brown

References

Brown, C., Arbour, J. and Jackson, D. 2012. Testing of the Effect of Missing Data Estimation and Distribution in Morphometric Multivariate Data Analyses. *Systematic Biology* 61(6):941-954.

See Also

missing.data,byclade,obliteratorGM

obliteratorGM	*Simulate missing geometric morphometric landmarks with anatomical bias*

Description

This is the geometric morphometric implementation of the LOST function obliterator. This attempts to replicate specimens showing regional incompleteness. For each specimen, this function randomly selects one starting data point for removal. All subsequent data points have a probability of removal that is proportional to the inverse of the distance to all previously removed data points, based on the shape of that particular specimen (this differs from the linear morphometric implementation which requires a reference set of coordinates). For a complete mathematical description see Brown et al. (2012).

Usage

```
obliteratorGM(x, remperc, expo=1)
```

Arguments

x	A n X m matrix of morphometric data with n specimens and m variables. Or a l X 2 or 3 X n array of geometric morphometric coordinates, with l being the number of landmarks.
remperc	The percentage of data to be removed from the matrix, expressed as a decimal (ex: 30 percent would be entered as 0.3)
expo	An optional term for raising the denominator to an exponent, to increase or decrease the severity of the anatomical bias

Value

Returns a n X m matrix of morphometric data with missing variables input as NA

Author(s)

J. Arbour and C. Brown

References

Brown, C., Arbour, J. and Jackson, D. 2012. Testing of the Effect of Missing Data Estimation and Distribution in Morphometric Multivariate Data Analyses. *Systematic Biology* 61(6):941-954.

See Also

missing.data,byclade,obliterator

Index