Package 'PELVIS'

June 25, 2020

Type Package

Title Probabilistic Sex Estimate using Logistic Regression, Based on
VISual Traits of the Human Os Coxae

Description
An R-Shiny application implementing a method of sexing the human os coxae based on logistic regressions and Bruzek's nonmetric traits <doi:10.1002/ajpa.23855>.

Version 2.0.2

Depends R (>= 4.0.0)

Imports DT, MASS, shiny

Suggests covr, testthat

License CeCILL-2 | file LICENSE

Encoding UTF-8

URL https://gitlab.com/f-santos/pelvis/

BugReports https://gitlab.com/f-santos/pelvis/issues/new

NeedsCompilation no

Author Frédéric Santos [aut, cre] (<https://orcid.org/0000-0003-1445-3871>),
Élodie Bernardeau [ctb]

Maintainer Frédéric Santos <frederic.santos@u-bordeaux.fr>

Repository CRAN

Date/Publication 2020-06-25 07:40:07 UTC

R topics documented:

1

PELVIS-package	*Probabilistic Sex Estimate using Logistic Regression, Based on VISual Traits of the Human Os Coxae*

Description

An R-Shiny application implementing a method of sexing the human os coxae based on logistic regressions and Bruzek's nonmetric traits.

Details

Package:	PELVIS
Type:	Package
Version:	2.0.2
Date:	2020-06-25
License:	CeCILL-2.1

Author(s)

Frédéric Santos, <frederic.santos@u-bordeaux.fr>

References

Bruzek, J. (2002) A method for visual determination of sex, using the human hip bone. *American Journal of Physical Anthropology* **117**, 157–168. doi: 10.1002/ajpa.10012

Santos, F., Guyomarc'h, P., Rmoutilova, R. and Bruzek, J. (2019) A method of sexing the human os coxae based on logistic regressions and Bruzek's nonmetric traits. *American Journal of Physical Anthropology* **169**(3), 435-447. doi: 10.1002/ajpa.23855

Examples

```
if(interactive()){ StartPELVIS() }
```

add_metavars	*Add the five Bruzek's main characters to a dataframe containing the eleven basic traits.*

Description

From a given dataset including the 11 visual traits exposed by Bruzek (2002), this function adds three corresponding main characters (PrSu, GrSN and InfP) based on the majority rule exposed in the original article.

Usage

```
add_metavars(dat)
```

Arguments

dat   A dataframe including the 11 visual traits described by Bruzek.

Value

A dataframe including also the main characters derived from those visual traits.

Note

This is mainly an internal function for the R-Shiny application implemented in PELVIS.

Author(s)

Frédéric Santos, <frederic.santos@u-bordeaux.fr>

References

Bruzek, J. (2002) A method for visual determination of sex, using the human hip bone. *American Journal of Physical Anthropology* **117**, 157–168. doi: 10.1002/ajpa.10012

Examples

```
# Load a dataset:
data(CTscanDataBruzek)
# Visualize the traits:
head(CTscanDataBruzek)
# Add all the main Bruzek's characters:
complete <- add_metavars(CTscanDataBruzek)
head(complete)
```

---

bruzek02   *Internal function for sexing the human os coxae using Bruzek's method (2002)*

Description

Produces a single (and non-probabilistic) sex estimate from five characters observed on the human os coxae, following Bruzek (2002)

Usage

```
bruzek02(x)
```

Arguments

x   A character vector of length 5, having three possible values: 'F', '0' or 'M'.

Value

One unique character value, 'F', 'I' or 'M', according to the majority rule exposed by Bruzek (2002).

Note

This is mainly an internal function for the R-Shiny application implemented in PELVIS.

Author(s)

Frédéric Santos, <frederic.santos@u-bordeaux.fr>

References

Bruzek, J. (2002) A method for visual determination of sex, using the human hip bone. *American Journal of Physical Anthropology* **117**, 157–168. doi: 10.1002/ajpa.10012

Examples

```
# Here we create manually an individual:
individual <- c(PrSu = "M", GrSN = "F",
                CArc = "F", InfP = "0", IsPu = "F")
individual
# Determination produced by Bruzek (2002): female individual.
bruzek02(individual)
```

CTscanDataBruzek   *Dataset including 198 virtually reconstructed ossa coxae*

Description

This dataset includes 198 ossa coxae segmented from CT-scans. The eleven trichotomic traits are given for each os coaxe (possibly with missing values for incomplete bones), along with the geographical origin and known sex of the individual. When possible, the age and stature of the individual are also given. This dataset is used as a training sample for the logistic regression models implemented in PELVIS.

Usage

```
data(CTscanDataBruzek)
```

Format

A data frame with 198 observations on the following 16 variables:

Id a factor with 198 levels (unique ID of each os coxae)

Indiv a factor with 99 levels (ID of each individual to whom the bone belongs)

Sex a factor with levels F, M (known sex)

Age a numeric vector (age of the associated individual in years)

Side a factor with levels L, R (left or right side)

PrSu1 an ordered factor with levels f, i, m

PrSu2 an ordered factor with levels f, i, m

PrSu3 an ordered factor with levels f, i, m

GrSN1 an ordered factor with levels f, i, m

GrSN2 an ordered factor with levels f, i, m

GrSN3 an ordered factor with levels f, i, m

CArc an ordered factor with levels F, 0, M

IsPu an ordered factor with levels F, 0, M

InfP1 an ordered factor with levels f, i, m

InfP2 an ordered factor with levels f, i, m

InfP3 an ordered factor with levels f, i, m

References

Santos, F., Guyomarc'h, P., Rmoutilova, R. and Bruzek, J. (2019) A method of sexing the human os coxae based on logistic regressions and Bruzek's nonmetric traits. *American Journal of Physical Anthropology* **169**(3), 435-447. doi: 10.1002/ajpa.23855

Bruzek, J., Rmoutilova, R., Guyomarc'h, P., & Santos, F. (2019) Supporting data for: A method of sexing the human os coxae based on logistic regressions and Bruzek's nonmetric traits [Data set]. Zenodo. http://doi.org/10.5281/zenodo.2589917

---

dataframe_sexing *Internal function for sexing several human ossa coxae using both original and revised Bruzek's methods (2002, 2019)*

Description

Produces sex estimates from each of the ossa coxae submitted by the user through the graphical user interface of the R-Shiny application.

Usage

```
dataframe_sexing(data, ref, updateProgressBar = NULL, conf_level = 0.95,
strategy = c("BIC", "AIC", "None"), trace = 1)
```

Arguments

data	A test dataset submitted by the user throught the graphical user interface. The predictive factors (i.e. the eleven trichotomic traits) should have the same headers and levels as in the reference dataset 'refData' included in PELVIS. An example of valid data file can be found on Zenodo: https://doi.org/10.5281/zenodo.2586897 (its field separator is the semicolon ";").
ref	A learning dataset for logistic regression models, basically the dataset 'refDataBruzek02' included in PELVIS (or any other dataset with the same variables).
updateProgressBar	
	Internal option for the R-Shiny application.
conf_level	0.95 by default, confidence level needed to produce a sex estimate.
strategy	A choice of information criterion ("BIC" or "AIC") for variable selection in logistic regression models, or "None" for no variable selection.
trace	See MASS::stepAIC.

Value

A complete dataframe of results displayed through the R-Shiny application.

Note

This is an internal function for the R-Shiny application implemented in PELVIS.

Author(s)

Frédéric Santos, <frederic.santos@u-bordeaux.fr>

References

Santos, F., Guyomarc'h, P., Rmoutilova, R. and Bruzek, J. (2019) A method of sexing the human os coxae based on logistic regressions and Bruzek's nonmetric traits. *American Journal of Physical Anthropology*, **169**(3), 435–447. doi: 10.1002/ajpa.23855

---

indiv_sexing	*Internal function for sexing one single human os coxae using revised Bruzek's method (2019)*

Description

Produces a statistical sex estimate from up to eleven characters observed on the human os coxae, following Bruzek (2019), and using logistic regression models.

Usage

```
indiv_sexing(ref, new_ind, strategy = c("BIC", "AIC", "None"), trace = 1,
conf_level = 0.95)
```

Arguments

ref	A learning dataset for logistic regression models, basically the dataset 'refData' included in PELVIS (or any other dataset with the same variables).
new_ind	A new os coxae to be determined, with eleven observed traits (possibly with missing values).
strategy	A choice of information criterion ("BIC" or "AIC") for variable selection in logistic regression models, or "None" for no variable selection.
trace	Passed to MASS::stepAIC.
conf_level	Required posterior probability threshold to produce a sex estimate.

Value

A list with the following components:

PredictedSex	One unique character value, 'F', 'I' or 'M': final sex estimate for the studied os coxae.
PostProb	Posterior probability for the individual to be a male.
BestModel	Best logistic regression model for the studied os coxae according to the BIC criterion.
VariablesUsed	Names of the variables (including part or all of the nonmissing traits for the studied os coaxe) used in this best model.
cvRate	Success rate in cross-validation. Cf. Santos et al. (2019) for more details about cross-validation here.
cvIndet	Rate of individuals remaining indeterminate using the best logistic regression model.

Note

This is mostly an internal function for the R-Shiny application implemented in PELVIS.

Author(s)

Frédéric Santos, <frederic.santos@u-bordeaux.fr>

References

Santos, F., Guyomarc'h, P., Rmoutilova, R. and Bruzek, J. (2019) A method of sexing the human os coxae based on logistic regressions and Bruzek's nonmetric traits. *American Journal of Physical Anthropology* **169**(3), 435-447. doi: 10.1002/ajpa.23855

Examples

```
data(refDataBruzek02)
# Pick the first individual of the reference dataset with its 11 traits, as an example:
individual <- refDataBruzek02[1, -c(1:6)]
individual
# Compute a sex estimate for this individual:
indiv_sexing(ref = refDataBruzek02, new_ind = individual)
```

---

refDataBruzek02	*Learning dataset for logistic regression models*

Description

This dataset includes 592 ossa coxae from five population samples. The eleven trichotomic traits are given for each os coaxe (possibly with missing values for incomplete bones), along with the geographical origin and known sex of the individual. When possible, the age and stature of the individual are also given. This dataset is used as a training sample for the logistic regression models implemented in PELVIS.

Usage

```
data(refDataBruzek02)
```

Format

A data frame with 592 observations on the following 17 variables:

Id a factor with 592 levels (unique ID of the individual to whom the bone belongs)

Orig a factor with 5 levels (geographical origin)

Sex a factor with levels F, M (known sex)

Age a numeric vector (age of the associated individual in years)

Side a factor with levels L, R (left or right side)

Stature a numeric vector (in cm)

PrSu1 an ordered factor with levels f, i, m

PrSu2 an ordered factor with levels f, i, m

PrSu3 an ordered factor with levels f, i, m

GrSN1 an ordered factor with levels f, i, m

GrSN2 an ordered factor with levels f, i, m

GrSN3 an ordered factor with levels f, i, m

CArc an ordered factor with levels F, 0, M

IsPu an ordered factor with levels F, 0, M

InfP1 an ordered factor with levels f, i, m

InfP2 an ordered factor with levels f, i, m

InfP3 an ordered factor with levels f, i, m

References

Santos, F., Guyomarc'h, P., Rmoutilova, R. and Bruzek, J. (2019) A method of sexing the human os coxae based on logistic regressions and Bruzek's nonmetric traits. *American Journal of Physical Anthropology* **169**(3), 435-447. doi: 10.1002/ajpa.23855

Bruzek, J., Rmoutilova, R., Guyomarc'h, P., & Santos, F. (2019) Supporting data for: A method of sexing the human os coxae based on logistic regressions and Bruzek's nonmetric traits [Data set]. Zenodo. http://doi.org/10.5281/zenodo.2589917

---

rightBonesDataBruzek   *Dataset including 518 right ossa coxae*

Description

This dataset includes 518 right ossa coxae from five population samples. These bones are used as a validation sample in the original study (Santos et al., 2019). The eleven trichotomic traits are given for each os coxae (possibly with missing values for incomplete bones), along with the geographical origin and known sex of the individual. When possible, the age and stature of the individual are also given.

Usage

```
data(rightBonesDataBruzek)
```

Format

A data frame with 518 observations on the following 17 variables:

Id a factor with 518 levels (unique ID of each os coxae)

Orig a factor with 5 levels (geographical origin or collection)

Sex a factor with levels F, M (known sex)

Age a numeric vector (age of the associated individual in years)

Side a factor with one single level R (right side)

Stature a numeric vector (in cm)

PrSu1 an ordered factor with levels f, i, m

PrSu2 an ordered factor with levels f, i, m

PrSu3 an ordered factor with levels f, i, m

GrSN1 an ordered factor with levels f, i, m

GrSN2 an ordered factor with levels f, i, m

GrSN3 an ordered factor with levels f, i, m

CArc an ordered factor with levels F, 0, M

IsPu an ordered factor with levels F, 0, M

InfP1 an ordered factor with levels f, i, m

InfP2 an ordered factor with levels f, i, m

InfP3 an ordered factor with levels f, i, m

References

Santos, F., Guyomarc'h, P., Rmoutilova, R. and Bruzek, J. (2019) A method of sexing the human os coxae based on logistic regressions and Bruzek's nonmetric traits. *American Journal of Physical Anthropology* **169**(3), 435-447. doi: 10.1002/ajpa.23855

Bruzek, J., Rmoutilova, R., Guyomarc'h, P., & Santos, F. (2019) Supporting data for: A method of sexing the human os coxae based on logistic regressions and Bruzek's nonmetric traits [Data set]. Zenodo. http://doi.org/10.5281/zenodo.2589917

---

start_pelvis   *An R-Shiny application for the sex estimation of the human os coxae*

Description

Launches a graphical user interface (GUI) allowing to use Bruzek's methods (2002, 2019) for sexing the human os coxae, based on eleven visual traits.

Usage

```
start_pelvis()
StartPELVIS()
```

Details

The R-Shiny application proposes two tabs:

- 'Data input: manual editing' can be used for both data entry and sex classification. The eleven trichotomic traits are manually edited for each os coxae through the GUI, and the corresponding sex estimates are then produced.

- 'Data input: from text file' is the classical way to get the sex estimates for a whole sample of ossa coxae correctly described in a file. PELVIS accepts .CSV or .TXT data files, but does not support .ODS or .XLS(X) files. The predictive factors (i.e. the eleven trichotomic traits) should have the same headers and levels as in the reference dataset 'refData' included in PELVIS. An example of valid data file can be found on Zenodo: `https://doi.org/10.5281/zenodo.2586897` (its field separator is the semicolon ";").

In both tabs, two sex estimates are given: the visual sex estimate from Bruzek (2002), and the probabilistic sex estimate from Santos, Guyomarc'h, Rmoutilova and Bruzek (2019). Depending on the traits possibly missing on the ossa coxae submitted to the program, the logistic regression models can use various subsets of best predictors (selected by AIC or BIC), or all predictors. The final subset of predictors used for each os coxae is given in the table of results. The user may also want to define a posterior probability threshold for sex estimation (0.90 or 0.95): any os coxae that does not reach this threshold will remain indeterminate.

Value

The function returns no value by itself, but the results can be downloaded through the graphical interface. The table of results includes the following columns:

- 'Sex estimate (Bruzek 2002)': the visual sex estimate based on Bruzek's method (2002).

- 'Statistical sex estimate (2019)': a sex estimation based on a logistic regression model, following the method described in Santos, Guyomarc'h, Rmoutilova and Bruzek (2019, submitted).

- 'Prob(M)' is the probability (obtained with the logistic regression model) that the individual is a man. According to tradition in biological anthropology, we have the following decsion rule: if Prob(M)>0.95 then the sex estimate is 'M'; if Prob(M)<0.05 then the sex estimate is 'F'; else the individual remains indeterminate ('I').

- 'Prob(F)', defined as 1-Prob(M), is the probability that the individual is a woman.
- 'Selected predictors in LR model': for a given individual, the sex estimation proceeds as follows. First, a complete model is built using all available (i.e., nonmissing) traits for this individual. Then, a classical stepwise model selection by BIC is performed, and the subset of the most useful traits is used to produce the final sex estimate. This column gives the traits used for each individual.
- '10-fold CV accuracy (%)': the rate of correct classification for the corresponding logistic regression model is estimated using a ten-fold cross-validation on the learning sample.
- 'Indet. rate in CV (%)': the rate of individuals remaining indeterminate in cross-validation for the corresponding logistic regression model.

Note

The R console is not available when the GUI is active. To exit the GUI, type Echap (on MS Windows systems) or Ctrl+C (on Linux systems) in the R console.

Regardless of the size and resolution of your screen, for convenience, it is advisable to decrease the zoom level of your web browser and/or to turn on fullscreen mode.

Author(s)

Frédéric Santos, <frederic.santos@u-bordeaux.fr>

References

Bruzek, J. (2002) A method for visual determination of sex, using the human hip bone. *American Journal of Physical Anthropology* **117**, 157–168. doi: 10.1002/ajpa.10012

Santos, F., Guyomarc'h, P., Rmoutilova, R. and Bruzek, J. (2019) A method of sexing the human os coxae based on logistic regressions and Bruzek's nonmetric traits. *American Journal of Physical Anthropology*, **169**(3), 435–447. doi: 10.1002/ajpa.23855

Examples

```
if(interactive()){start_pelvis()}
```

Index