Package ‘TestDimorph’

April 23, 2020

Type Package

Title Analysis Of The Interpopulation Difference In Degree of Sexual Dimorphism Using Summary Statistics

Version 0.3.1

Maintainer Bassam A. Abulnoor `<bas12@fayoum.edu.eg>`

Description Provides two approaches of comparison; the univariate and the multivariate analysis in two or more populations. Since the main obstacle of performing systematic comparisons in anthropological studies is the absence of raw data, the current package offer a solution for this problem by allowing the use of published summary statistics of metric data (mean, standard deviation and sex specific sample size) as illustrated by the works of Greene, D. L. (1989) <doi:10.1002/ajpa.1330790113> and Konigsberg, L. W. (1991) <doi:10.1002/ajpa.1330840110>.

Imports
Rfast,plyr,stats,utils,reshape2,purrr,dplyr,caret,rlang,MASS,klaR,corrplot,truncnorm,stringr,ggplot2,plotROC,DescTools,t

Suggests testthat (>= 2.1.0), AnthropMMD, Rdpack

Depends R (>= 2.10)

RdMacros Rdpack

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.1.0

Language en-US

URL https://github.com/bassam-abulnoor/TestDimorph

BugReports https://github.com/bassam-abulnoor/TestDimorph/issues

NeedsCompilation no

Author Bassam A. Abulnoor [aut, cre] (<https://orcid.org/0000-0003-4351-2754>), MennattAllah H. Attia [aut] (<https://orcid.org/0000-0002-2304-532X>), Lyle W. Konigsberg [ctb, dtc] (<https://orcid.org/0000-0003-4052-1575>)

Repository CRAN

Date/Publication 2020-04-23 12:40:02 UTC

R topics documented:

`AccuModel` *Evaluation Of Sex-prediction Accuracy*

Description

Testing and visualization of the accuracy of different sex prediction models using the confusionMatrix and roc curves

Usage

```
AccuModel(
  f,
  x,
  y,
  Sex = 1,
  Pop = 2,
  byPop = TRUE,
  method = "lda",
  plot = FALSE,
  cutoff = 0.5,
  ref. = "F",
  post. = "M",
  ...
)
```

Arguments

`f`	Formula in the form `groups ~ x1 + x2 + ....` The grouping factor is placed to the left hand side while the numerical measurements are placed to the right hand side
`x`	Data frame to be fitted to the model
`y`	New data frame to be tested

`Sex`	Number of the column containing sex 'M' for male and 'F' for female, Default: 1
`Pop`	Number of the column containing populations' names, Default: 2
`byPop`	Logical; if TRUE returns the accuracy in different populations of the new data frame, Default: TRUE.
`method`	Different methods of modeling see `details` , Default:'lda'
`plot`	Logical; if TRUE returns an roc curve for model accuracy, Default: FALSE
`cutoff`	cutoff value when using logistic regression, Default: 0.5
`ref.`	reference category in the grouping factor, Default: 'F'
`post.`	positive category in the grouping factor, Default: 'M'
`...`	additional arguments that can passed to modeling, confusionMatrix function and roc curve generated by geom_roc

Details

Tibble/data frames to be entered as input need to be arranged in a similar manner to Howells dataset. Methods used for modeling are:

lda	linear discriminant analysis
qda	quadratic discriminant analysis
rda	regularized discriminant analysis
glm	binomial logistic regression
raf	random forest

Value

Visual and numerical accuracy parameters for the tested model

See Also

`lda`,`qda` `rdarandomForest` `GeomRoc` `confusionMatrix`

Examples

```
#Splitting Howells dataset into training and test datasets
smp_size <- floor(0.5 * nrow(Howells))
set.seed(123)
train_ind <- sample(seq_len(nrow(Howells)), size = smp_size)
train <- Howells[train_ind, ]
test <- Howells[-train_ind, ]
library(TestDimorph)
AccuModel(
Sex ~ GOL + NOL + BNL,
x = train,
y = test,
byPop = FALSE,
method = "lda",
plot = FALSE
)
```

`aovSS` *Sex-Specific One-way ANOVA From Summary statistics*

Description

Calculates sex specific one-way ANOVA followed by from summary statistics.

Usage

```
aovSS(
  x,
  Pop = 1,
  pairwise = TRUE,
  letters = FALSE,
  es = FALSE,
  digits = 4,
  method = "hsd",
  sig.level = 0.05
)
```

Arguments

`x`	Tibble/data frame containing summary statistics, Default: NULL
`Pop`	Number of the column containing populations' names, Default: 1
`pairwise`	Logical; if TRUE runs multiple pairwise comparisons on different populations using post hoc test of choice, Default: TRUE
`letters`	Logical; if TRUE returns letters for pairwise comparisons where significantly different populations are given different letters, Default: FALSE'
`es`	Logical; if TRUE effect size is included in the output , Default: FALSE
`digits`	Number of significant digits, Default: 4
`method`	Type of post hoc test implemented by PostHocTest, Default: 'hsd'
`sig.level`	Critical p.value, Default: 0.05

Details

Data is entered as a tibble/data frame of summary statistics where the column containing population names is chosen by position (first by default), other columns of summary data should have specific names (case sensitive) similar to baboon.parms_df

Value

Sex specific ANOVA tables and pairwise comparisons in tidy format.

See Also

`PostHocTest`

Examples

```
  # Comparisons of femur head diameter in four populations
  library(TestDimorph)
m <- c(150.00, 82.00, 36.00, 34.00)
f <- c(150.00, 58.00, 34.00, 24.00)
M.mu <- c(49.39, 48.33, 46.99, 45.20)
F.mu <- c(42.91, 42.89, 42.44, 40.90)
M.sdev <- c(3.01, 2.53, 2.47, 2.00)
F.sdev <- c(2.90, 2.84, 2.26, 2.90)
df <- cbind.data.frame(
  Pop = c('Turkish', 'Bulgarian', 'Greek', 'Portuguese '),
  m,
  f,
  M.mu,
  F.mu,
  M.sdev,
  F.sdev,
  stringsAsFactors = TRUE
)
aovSS(x = df)
```

`baboon.parms_df` *Summary statistics of baboon data collection-list*

Description

A dataset containing summary statistics for low density lipoprotein (LDL) and apolipoprotein-B (apo-B) levels in 604 baboons measured on two different diets: a basal diet ’chow’ and a high cholesterol, saturated fat diet ’pink’ (HCSF). The baboons were classified into one of three subspecies (Papio hamadryas anubis, P.h. cynocephalus, or anubistcynocephalus hybrid). Each animal was measured on each of the two diets.

Usage

`baboon.parms_df`

Format

A data frame with 12 rows and 8 variables

Trait Type of apolipoprotein

Sub Type of species

M.mu Means of lipoproteins in different species for males

F.mu Means of lipoproteins in different species for females

m Male sample sizes

f Female sample sizes

M.sdev Standard deviations for males

F.sdev Standard deviations for females

Note

The baboon data collection were supported by NIH grant HL28972 and NIH contract HV53030 to the Southwest Foundation for Biomedical Research (Now: Texas Biomedical Research Institute), and funds from the Southwest Foundation for Biomedical Research

References

Konigsberg LW (1991). “An historical note on the t-test for differences in sexual dimorphism between populations.” *American journal of physical anthropology*, **84**(1), 93–96.

`baboon.parms_list` *Summary statistics of baboon data collection-data frame*

Description

A dataset containing summary statistics for low density lipoprotein (LDL) and apolipoprotein-B (apo-B) levels in 604 baboons measured on two different diets: a basal diet ’chow’ and a high cholesterol, saturated fat diet ’pink’ (HCSF). The baboons were classified into one of three subspecies (Papio hamadryas anubis, P.h. cynocephalus, or anubistcynocephalus hybrid). Each animal was measured on each of the two diets.

Usage

```
baboon.parms_list
```

Format

A list of 7 matrices.

R.res pooled within group correlation matrix

M.mu Means of lipoproteins in different species for males

F.mu Means of lipoproteins in different species for females

m Male sample sizes

f Female sample sizes

M.sdev Standard deviations for males

F.sdev Standard deviations for females

Note

The baboon data collection were supported by NIH grant HL28972 and NIH contract HV53030 to the Southwest Foundation for Biomedical Research (Now: Texas Biomedical Research Institute), and funds from the Southwest Foundation for Biomedical Research

References

Konigsberg LW (1991). “An historical note on the t-test for differences in sexual dimorphism between populations.” *American journal of physical anthropology*, **84**(1), 93–96.

`extract_sum` *Summary Statistics Extraction*

Description

Extract summary data needed for other functions from raw data.

Usage

```
extract_sum(x, Sex = 1, Pop = 2, firstX = 3, test = 1, run = TRUE, ...)
```

Arguments

`x`	Tibble/data frame containing raw data.
`Sex`	Number of the column containing sex 'M' for male and 'F' for female, Default: 1
`Pop`	Number of the column containing populations' names, Default: 2
`firstX`	Number of column containing measured parameters (First of multiple in case of multivariate analysis), Default: 3
`test`	`1` for Greene t-test Tg, `2` for univariate, `3` for sex specific ANOVA aovSS, and `4` for multivariate, Default: 1
`run`	Logical; if TRUE runs the corresponding test after data extraction, Default: TRUE
`...`	Additional arguments that could be passed to the test of choice

Details

Raw data is entered in a wide format tibble/data frame similar to Howells data set. The first two columns contain sex `Sex` (`M` for male and `F` for female) (Default: `1`) and populations' names `Pop` (Default: `2`). Starting from `firstX` column (Default: `3`), measured parameters are entered each in a separate column.

Value

Input for other functions.

Examples

```
# for multivariate test
library(TestDimorph)
extract_sum(Howells,test=4)
# for univariate test on a specific parameter
library(TestDimorph)
extract_sum(Howells, test = 2,firstX = 4)
```

`Howells` *The Howells' craniometric data*

Description

A subset of a dataset that consists of 82 craniometric measurements taken from approximately two thousands and half human crania from 28 geographically diverse populations.

Usage

```
Howells
```

Format

A data frame with 441 rows and 10 variables:

Sex 'M' for male and 'F' for female

Pop Populations' names

GOL Glabello occipital length

NOL Nasio occipital length

BNL Bastion nasion length

BBH Basion bregma height

XCB Maximum cranial breadth

XFB Max frontal breadth

ZYB Bizygomatic breadth

AUB Biauricular breadth

References

Howells WW (1995). "Who's who in skulls: ethnic identification of crania from measurements." *Papers of the Peabody Museum of Archaeology and Ethnology*, **82**.

Howells WW (1989). "Skull shapes and the map: craniometric analyses in the dispersion of modern Homo." *Papers of the Peabody Museum of Archaeology and Ethnology*, **79**.

Howells WW (1996). "Howells' craniometric data on the internet." *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, **101**(3), 441–442.

Howells WW (1973). "Cranial variation in man: a study by multivariate analysis of patterns of difference among recent human populations." *Peabody Museum of Archaeology and Ethnology, Harvard Univ.*.

`multivariate` *Multivariate Analysis Of Sexual Dimorphism*

Description

Multivariate extension of Greene t-test Tg

Usage

```
multivariate(
  x,
  R.res = NULL,
  Parms = 1,
  Pop = 2,
  es = FALSE,
  univariate = FALSE,
  padjust = "none",
  ...,
  lower.tail = FALSE,
  digits = 4
)
```

Arguments

`x`	Tibble/Data frame or list containing summary statistics for multiple parameters measured in both sexes in two or more populations.
`R.res`	Pooled within correlational matrix, Default: NULL
`Parms`	Number of the column containing names of measured parameters, Default: 1
`Pop`	Number of the column containing populations' names, Default: 2
`es`	Logical; if TRUE effect size is included in the output , Default: FALSE
`univariate`	Logical; if TRUE conducts multiple univariate analyses on different parameters separately, Default: FALSE
`padjust`	Method of p.value adjustment for multiple comparisons following p.adjust.methods, Default: 'none'
`...`	Additional arguments that could be passed to the univariate function
`lower.tail`	Logical; if TRUE probabilities are `P[X <= x]`, otherwise, `P[X > x]`., Default: FALSE
`digits`	Number of significant digits, Default: 4

Details

Data can be entered either as a tibble/data frame of summary statistics as in baboon.parms_df . In that case the pooled within correlational matrix `R.res` should be entered as a separate argument as in R. Another acceptable format is a named list of matrices containing different summary statistics as well as the correlational matrix as in baboon.parms_list. By setting the option `univariate` to `TRUE`, multiple `ANOVAs` can be run on each parameter independently with the required p value correction using p.adjust.methods.

Value

Tibble of MANOVA results

References

Konigsberg LW (1991). "An historical note on the t-test for differences in sexual dimorphism between populations." *American journal of physical anthropology*, **84**(1), 93–96.

Examples

```
# x is a data frame with separate correlational matrix
library(TestDimorph)
multivariate(baboon.parms_df, R.res = R)
# x is a list with the correlational matrix included
library(TestDimorph)
multivariate(baboon.parms_list, univariate = TRUE, padjust = 'bonferroni')
```

`R` *Pooled within group correlation matrix for baboon data*

Description

Pooled within group correlation matrix for baboon data

Usage

```
R
```

Format

A 4*4 numerical matrix

`RawGen` *Raw Data Generation By Log-normal Or Truncated Distribution*

Description

Generates raw data from summary statistics using uni/multivariate log/truncated normal distribution

Usage

```
RawGen(
  x,
  Parms = 1,
  Pop = 2,
  R.res = NULL,
  dist = "trunc",
  lower = -Inf,
  upper = Inf,
  format = "wide",
  complete_cases = FALSE
)
```

Arguments

`x`	Tibble/Data frame or list containing summary statistics for multiple parameters measured in both sexes in two or more populations.
`Parms`	Number of the column containing names of measured parameters, Default: 1
`Pop`	Number of the column containing populations' names, Default: 2
`R.res`	Pooled within correlational matrix, Default: NULL
`dist`	univariate distribution used for data generation either `log` for log-normal or `trunc` for truncated, Default: 'trunc'
`lower`	vector of lower bounds, Default: -Inf
`upper`	vector of upper bounds, Default: Inf
`format`	form of the resultant tibble either 'long' or 'wide', Default: 'wide'
`complete_cases`	Logical; if TRUE rows with missing values will be removed, Default: FALSE

Details

If data generation is desired using multivariate distribution data is entered in the form of a list of summary statistics and pooled within correlational matrix as in baboon.parms_list, or the summary statistics are entered separately in the form of a data frame/tibble as in baboon.parms_df with a separate correlational matrix as in R. If data frame/tibble is entered without a correlational matrix, data generation is carried out using univariate distribution. N.B: Transformation of raw summary data to logged data is only possible for univariate distribution and if multivariate log-normal distribution is desired logged values should be entered directly with `dist` set to `trunc`.

Value

tibble of raw data

References

Hussein MHA, Abulnoor BAE (2019). “Sex estimation of femur using simulated metapopulation Database: a preliminary investigation.” *Forensic Science International: Reports*. ISSN 2665-9107, doi: 10.1016/j.fsir.2019.100009.

Examples

```
# Data generation using univariate distribution
library(TestDimorph)
RawGen(baboon.parms_df)
# Data generation using multivariate distribution
library(TestDimorph)
RawGen(baboon.parms_list)
```

`Tg` *Greene t-test of Sexual Dimorphism*

Description

Calculation and visualization of the differences in degree sexual dimorphism between two populations using summary statistics as input.

Usage

```
Tg(
  x = NULL,
  Pop = 1,
  es = FALSE,
  plot = FALSE,
  ...,
  alternative = "two.sided",
  padjust = "none",
  letters = FALSE,
  digits = 4,
  sig.level = 0.05,
  N = NULL,
  m = NULL,
  m2 = NULL,
  f = NULL,
  f2 = NULL,
  M.mu = NULL,
  M.mu2 = NULL,
  F.mu = NULL,
```

```
  F.mu2 = NULL,
  M.sdev = NULL,
  M.sdev2 = NULL,
  F.sdev = NULL,
  F.sdev2 = NULL
)
```

Arguments

`x`	Tibble/data frame containing summary statistics, Default: NULL
`Pop`	Number of the column containing populations' names, Default: 1
`es`	Logical; if TRUE effect size is included in the output , Default: FALSE
`plot`	Logical; if TRUE graphical matrix of p-values, Default: TRUE
`...`	additional arguments that can be passed to corrplot function.
`alternative`	a character string specifying the alternative hypothesis, must be one of "two.sided", "greater" or "less", Default: 'two.sided'
`padjust`	Method of p.value adjustment for multiple comparisons following p.adjust.methods, Default: 'none'
`letters`	Logical; if TRUE returns letters for pairwise comparisons where significantly different populations are given different letters, Default: FALSE'
`digits`	Number of significant digits, Default: 4
`sig.level`	Critical p.value, Default: 0.05
`N`	Number of pairwise comparisons for p.adjust.methods, if left `NULL` it will follow the formula `n(n-1)/2` where `n` is the number of populations , Default: NULL
`m`	Number of male sample size in the first population, Default: NULL
`m2`	Number of male sample size in the second population, Default: NULL
`f`	Number of female sample size in the first population, Default: NULL
`f2`	Number of female sample size in the second population, Default: NULL
`M.mu`	Means for males in the first population, Default: NULL
`M.mu2`	Means for males in the second population, Default: NULL
`F.mu`	Means for females in the first population, Default: NULL
`F.mu2`	Means for females in the second population, Default: NULL
`M.sdev`	Standard deviation for males in the first population, Default: NULL
`M.sdev2`	Standard deviation for males in the second population, Default: NULL
`F.sdev`	Standard deviation for females in the first population, Default: NULL
`F.sdev2`	Standard deviation for females in the second population, Default: NULL

Details

Summary statistics can be entered directly as arguments in case of comparing two populations or as a tibble/data frame of summary statistics where the column containing population names is chosen by position (first by default), other columns of summary data should have specific names (case sensitive) similar to baboon.parms_df

Value

Tibble of t.test results

References

Greene DL (1989). "Comparison of t-tests for differences in sexual dimorphism between populations." *American Journal of Physical Anthropology*, **79**(1), 121–125. Timonov P, Fasova A, Radoinova D, Alexandrov A, Delev D (2014). "A study of sexual dimorphism in the femur among contemporary Bulgarian population." *Eurasian Journal of Anthropology*, **5**(2), 46–53. Gulhan O, Harrison K, Kiris A (2015). "A new computer-tomography-based method of sex estimation: Development of Turkish population-specific standards." *Forensic science international*, **255**, 2–8.

See Also

`multcompLetters` `corrplot`

Examples

```
#Comparisons of femur head diameter in four populations
library(TestDimorph)
m <- c(150.00, 82.00, 36.00, 34.00)
f <- c(150.00, 58.00, 34.00, 24.00)
M.mu <- c(49.39, 48.33, 46.99, 45.20)
F.mu <- c(42.91, 42.89, 42.44, 40.90)
M.sdev <- c(3.01, 2.53, 2.47, 2.00)
F.sdev <- c(2.90, 2.84, 2.26, 2.90)
df <- cbind.data.frame(
  Pop = c('Turkish', 'Bulgarian', 'Greek', 'Portuguese '),
  m,
  f,
  M.mu,
  F.mu,
  M.sdev,
  F.sdev,
  stringsAsFactors = TRUE
)
Tg(
   df,
   plot = TRUE,
   method = 'ellipse',
  type = 'lower',
   col = c(
       '#AEB6E5',
       '#B1A0DB',
       '#B788CD',
       '#BC6EB9',
       '#BC569E',
       '#B6407D',
       '#A93154'
   ),
   tl.cex = 0.8,
```

```
    tl.col = 'black',
   insig =
        'label_sig',
    tl.srt = 0.1,
    pch.cex = 2.5,
    tl.pos = 'ld',
    win.asp = 1,
    number.cex = 0.5,
    na.label = 'NA'
 )
```

`univariate` *Univariate Analysis Of Sexual Dimorphism*

Description

Calculation and visualization of the differences in degree sexual dimorphism between multiple populations using a modified one-way ANOVA and summary statistics as input

Usage

```
univariate(
  x,
  Pop = 1,
  es = FALSE,
  pairwise = FALSE,
  padjust = "none",
  ...,
  lower.tail = FALSE,
  N = NULL,
  digits = 4
)
```

Arguments

`x`	Tibble/data frame containing summary statistics, Default: NULL
`Pop`	Number of the column containing populations' names, Default: 1
`es`	Logical; if TRUE effect size is included in the output , Default: FALSE
`pairwise`	Logical; if TRUE runs multiple pairwise comparisons on different populations using Tg test, Default: FALSE
`padjust`	Method of p.value adjustment for multiple comparisons following p.adjust.methods, Default: 'none'
`...`	Additional arguments that could be passed to the Tg function
`lower.tail`	Logical; if TRUE probabilities are `P[X <= x]`, otherwise, `P[X > x]`., Default: FALSE
`N`	Number of pairwise comparisons for p.adjust.methods, if left `NULL` it will follow the formula `n(n-1)/2` where `n` is the number of populations , Default: NULL
`digits`	Number of significant digits, Default: 4

Details

Data is entered as a tibble/data frame of summary statistics where the column containing population names is chosen by position (first by default), other columns of summary data should have specific names (case sensitive) similar to baboon.parms_df

Value

Tibble of ANOVA results

References

Konigsberg LW (1991). “An historical note on the t-test for differences in sexual dimorphism between populations.” *American journal of physical anthropology*, **84**(1), 93–96. Timonov P, Fasova A, Radoinova D, Alexandrov A, Delev D (2014). “A study of sexual dimorphism in the femur among contemporary Bulgarian population.” *Eurasian Journal of Anthropology*, **5**(2), 46–53. Curate F, Umbelino C, Perinha A, Nogueira C, Silva AM, Cunha E (2017). “Sex determination from the femur in Portuguese populations with classical and machine-learning classifiers.” *Journal of forensic and legal medicine*, **52**, 75–81. Kranioti EF, Vorniotakis N, Galiatsou C, Iscan MY, Michalodimitrakis M (2009). “Sex identification and software development using digital femoral head radiographs.” *Forensic science international*, **189**(1-3), 113–e1. Gulhan O, Harrison K, Kiris A (2015). “A new computer-tomography-based method of sex estimation: Development of Turkish population-specific standards.” *Forensic science international*, **255**, 2–8.

Examples

```
# Comparisons of femur head diameter in four populations
library(TestDimorph)
m <- c(150.00, 82.00, 36.00, 34.00)
f <- c(150.00, 58.00, 34.00, 24.00)
M.mu <- c(49.39, 48.33, 46.99, 45.20)
F.mu <- c(42.91, 42.89, 42.44, 40.90)
M.sdev <- c(3.01, 2.53, 2.47, 2.00)
F.sdev <- c(2.90, 2.84, 2.26, 2.90)
df <-
cbind.data.frame(
   Pop = c('Turkish', 'Bulgarian', 'Greek', 'Portuguese '),
   m,
   f,
   M.mu,
   F.mu,
   M.sdev,
   F.sdev,
   stringsAsFactors = TRUE
)
univariate(df, pairwise = TRUE, padjust = 'bonferroni')
```

Index