

Package ‘TestDimorph’

September 22, 2020

Type Package

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

Version 0.3.4

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>.

License GPL-3

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

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

Depends R (>= 2.10)

Imports caret, corrplot, cutpointr, dplyr, ggplot2, multcompView, Rfast, stats, tibble, tidyr, tmvtnorm, truncnorm, utils

Suggests AnthropMMD, testthat (>= 2.1.0), vdiffr

Encoding UTF-8

Language en-US

LazyData true

RoxygenNote 7.1.1

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 [aut] (<<https://orcid.org/0000-0003-4052-1575>>)

Repository CRAN

Date/Publication 2020-09-22 06:30:19 UTC

R topics documented:

accu_model	2
aov_ss	4
baboon.parms_df	5
baboon.parms_list	6
extract_sum	7
Howells	8
multivariate	9
R	10
raw_gen	10
t_greene	11
univariate	13
Index	16

accu_model	<i>Evaluation Of Sex prediction Accuracy</i>
------------	--

Description

Testing, cross validation and visualization of the accuracy of different sex prediction models using the [confusionMatrix](#) and roc curves

Usage

```
accu_model(  
  f,  
  x,  
  y = NULL,  
  method = "lda",  
  res_method = "repeatedcv",  
  p = 0.75,  
  nf = 10,  
  nr = 3,  
  plot = FALSE,  
  Sex = 1,  
  Pop = NULL,  
  byPop = FALSE,  
  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, if NULL x is splitted to test and training data set, Default: NULL
method	A string specifying which classification or regression model to use,
res_method	The resampling method used by trainControl , Default: 'repeatedcv'
p	Percentage of x for testing the model in case y is NULL, Default: 0.75
nf	number of folds or of resampling iterations, Default: 10
nr	Number of repeats for repeated k fold cross validation, Default: 3
plot	Logical; if TRUE returns an roc curve for model accuracy, Default: FALSE
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: NULL
byPop	Logical; if TRUE returns the accuracy in different populations of the new data frame, Default: FALSE.
ref.	reference category in the grouping factor, Default: 'F'
post.	positive category in the grouping factor, Default: 'M'
...	additional arguments that can be passed to modeling, confusionMatrix function and roc curve generated by plot_roc

Details

Data frames to be entered as input need to be arranged in a similar manner to [Howells](#) dataset.

Value

Visual and numerical accuracy parameters for the tested model

See Also

[cutpointr::plot_roc\(\)](#) [caret::confusionMatrix\(\)](#)

Examples

```
# Using a single dataset
library(TestDimorph)
accu_model(
  Sex ~ GOL + NOL + BNL,
  x = Howells,
  method = "lda",
  plot = FALSE
)
```

aov_ss

*Sex Specific One way ANOVA From Summary statistics***Description**

Calculates sex specific one way ANOVA from summary statistics.

Usage

```
aov_ss(
  x,
  Pop = 1,
  pairwise = TRUE,
  letters = FALSE,
  es = FALSE,
  digits = 4,
  sig.level = 0.05
)
```

Arguments

x	A data frame containing summary statistics.
Pop	Number of the column containing populations' names, Default: 1
pairwise	Logical; if TRUE runs multiple pairwise comparisons on different populations using Tukey's post hoc test, 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
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.params_df](#)

Value

Sex specific ANOVA tables and pairwise comparisons in tidy format.

Examples

```
# Comparisons of femur head diameter in four populations
library(TestDimorph)
df <- data.frame(
  Pop = c("Turkish", "Bulgarian", "Greek", "Portuguese "),
  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)
)
aov_ss(x = df)
```

baboon.parms_df	<i>Summary statistics of baboon data collection list 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 anubiscynocephalus hybrid). Each animal was measured on each of the two diets.</i>
-----------------	---

Description

Summary statistics of baboon data collection list 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 anubiscynocephalus 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	<i>Summary statistics of baboon data collection data frame 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 anubiscynocephalus hybrid). Each animal was measured on each of the two diets.</i>
-------------------	---

Description

Summary statistics of baboon data collection data frame 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 anubiscynocephalus 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	<i>Summary Statistics Extraction</i>
-------------	--------------------------------------

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	Data frame of 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 t_greene , 2 for univariate , 3 for sex specific ANOVA aov_ss , 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 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 (1989). *Skull shapes and the map: craniometric analyses in the dispersion of modern Homo..* Papers of the Peabody Museum of Archaeology and Ethnology, 79.

Description

Multivariate extension of Greene t test [t_greene](#)

Usage

```
multivariate(
  x,
  R.res = NULL,
  Parms = 1,
  Pop = 2,
  es = FALSE,
  univariate = FALSE,
  padjust = p.adjust.methods,
  ...,
  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
...	Additional arguments that could be passed to the univariate function
lower.tail	Logical; if TRUE probabilities are $P[X \leq x]$, 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

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	<i>Pooled within group correlation matrix for baboon data</i>
---	---

Description

Pooled within group correlation matrix for baboon data

Usage

R

Format

A 4*4 numerical matrix

raw_gen	<i>Raw Data Generation By Log-normal Or Truncated Distribution</i>
---------	--

Description

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

Usage

```
raw_gen(
  x,
  Parm = 1,
  Pop = 2,
  R.res = NULL,
  dist = c("truncated", "log"),
  lower = -Inf,
  upper = Inf,
  format = c("wide", "long"),
  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 truncated for truncated, Default: 'truncated'
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 truncated.

Value

tibble of raw data

Examples

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

t_greene

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

```
t_greene(
  x,
  Pop = 1,
  es = FALSE,
  plot = FALSE,
  ...,
  alternative = c("two.sided", "less", "greater"),
  padjust = p.adjust.methods,
  letters = FALSE,
  digits = 4,
  sig.level = 0.05
)
```

Arguments

x	A data frame containing summary statistics.
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: FALSE
...	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".
padjust	Method of p.value adjustment for multiple comparisons following p.adjust.methods
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

Details

The input is a 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

See Also

[multcompView::multcompLetters\(\)](#) [corrplot::corrplot\(\)](#)

Examples

```
# Comparisons of femur head diameter in four populations
library(TestDimorph)
df <- data.frame(
  Pop = c("Turkish", "Bulgarian", "Greek", "Portuguese "),
  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)
)
t_greene(
  df,
  plot = TRUE,
  method = "ellipse",
  padjust = "none",
  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"
)
```

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 = p.adjust.methods,
  ...,
  lower.tail = FALSE,
  N = NULL,
  digits = 4
)
```

Arguments

x	A data frame containing summary statistics.
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 t_greene test, Default: FALSE
padjust	Method of p.value adjustment for multiple comparisons following p.adjust.methods
...	Additional arguments that could be passed to the t_greene function
lower.tail	Logical; if TRUE probabilities are $P[X \leq x]$, 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

Examples

```
# Comparisons of femur head diameter in four populations
library(TestDimorph)
df <-
  data.frame(
    Pop = c("Turkish", "Bulgarian", "Greek", "Portuguese "),
    m = c(150.00, 82.00, 36.00, 34.00),
```

```
M.mu = c(49.39, 48.33, 46.99, 45.20),  
M.sdev = c(3.01, 2.53, 2.47, 2.00),  
f = c(150.00, 58.00, 34.00, 24.00),  
F.mu = c(42.91, 42.89, 42.44, 40.90),  
F.sdev = c(2.90, 2.84, 2.26, 2.90)  
)  
univariate(df, pairwise = TRUE, padjust = "bonferroni")
```

Index

* datasets

- `baboon.parms_df`, 5
- `baboon.parms_list`, 6
- `Howells`, 8
- `R`, 10

`accu_model`, 2

`aov_ss`, 4, 7

`baboon.parms_df`, 4, 5, 9, 11, 12, 14

`baboon.parms_list`, 6, 9, 11

`caret::confusionMatrix()`, 3

`confusionMatrix`, 2, 3

`corrplot`, 12

`corrplot::corrplot()`, 12

`cutpointr::plot_roc()`, 3

`extract_sum`, 7

`Howells`, 3, 7, 8

`multcompView::multcompLetters()`, 12

`multivariate`, 7, 9

`p.adjust.methods`, 9, 12, 14

`plot_roc`, 3

`R`, 9, 10, 11

`raw_gen`, 10

`t_greene`, 7, 9, 11, 14

`trainControl`, 3

`univariate`, 7, 9, 13