Package ‘effsize’

April 22, 2015

Type Package
Title Efficient Effect Size Computation
Version 0.5.4
Date 2015-04-21
Description A collection of functions to compute the standardized effect sizes for experiments (Cohen d, Hedges g, Cliff delta, Vargha-Delaney A). The computation algorithms have been optimized to allow efficient computation even with very large data sets.

URL http://github.com/mtorchiano/effsize/

BugReports https://github.com/mtorchiano/effsize/issues
License GPL-2
NeedsCompilation no
Repository CRAN
Author Marco Torchiano [aut, cre]
Maintainer Marco Torchiano <marco.torchiano@polito.it>
Date/Publication 2015-04-22 00:05:53

R topics documented:

effsize-package .................................................. 2
cliff.delta ......................................................... 3
cohen.d ........................................................... 5
print.effsize ..................................................... 7
VD.A ............................................................... 8

Index 10
Description

This package contains functions to compute effect sizes both based on means difference (Cohen’s d and Hedges g), dominance matrices (Cliff’s Delta) and stochastic superiority (Vargha-Delaney A).

The computation (especially for Cliff’s Delta) is carried on with highly efficient algorithms.

Details

The main functions are:

- `cliff.delta`
- `cohen.d`
- `vdNa`

Change history

0.3.1 Fixed a bug in `cohen.d` when PAIRED=TRUE, now the PAIRED parameter has no effect, it is left just for compatibility. In a future code clean-up it may be removed.

0.4 Implemented a new algorithm with improved memory and time complexity. In particular new time complexity is \( T = O(n1*\log(n2)) \) vs. the previous \( T = O(n1*n2) \), and new memory complexity \( M = O(n1 + n2) \) vs. the previous \( M = O(n1 * n2) \). In practice now the computation becomes feasible in a "reasonable" time.

0.4.1 Code clean-up and optimization using vectorized binary partitioning.

0.5 Added Vargha and Delaney A and fixed minor bugs with Cohen.d.

0.5.1 Modified the Vargha and Delaney A computation to minimize accuracy errors.

0.5.2 Fixed bug in `cliff.delta`.

0.5.3 Fixed bug in `cohen.d.formula`.

0.5.4 Fixed minor issue detected by check.

Author(s)

Marco Torchiano [http://softeng.polito.it/torchiano/](http://softeng.polito.it/torchiano/)
cliff.delta

Cliff’s Delta effect size for ordinal variables

Description

Computes the Cliff’s Delta effect size for ordinal variables with the related confidence interval using efficient algorithms.

Usage

cliff.delta(treatment, ... )

## S3 method for class 'formula'
cliff.delta(formula, data=list(), conf.level=.95, 
use.unbiased=TRUE, use.normal=FALSE, 
return.dm=FALSE, ...)

## Default S3 method:
cliff.delta(treatment, control, conf.level=.95, 
use.unbiased=TRUE, use.normal=FALSE, 
return.dm=FALSE, ...)

Arguments

treatment numeric vector or ordered factor of data values for the treatment group (see Details)
control numeric vector or ordered factor of data values for the control group (see Details)
conf.level confidence level of the confidence interval
use.unbiased a logical indicating whether to compute the delta’s variance using the "unbiased" estimate formula or the "consistent" estimate
use.normal logical indicating whether to use the normal or Student-t distribution for the confidence interval estimation
return.dm logical indicating whether to return the dominance matrix. **Warning**: the explicit computation of the dominance uses a sub-optimal algorithm both in terms of memory and time
formula a formula of the form \( y \sim f \), where \( y \) is a numeric variable giving the data values and \( f \) a factor with two levels giving the corresponding group
data an optional matrix or data frame containing the variables in the formula formula. By default the variables are taken from environment(formula).
... further arguments to be passed to or from methods.
Details

Uses the original formula reported in (Cliff 1996).

If the dominance matrix is required i.e. return.dm=TRUE) the full matrix is computed thus using the naive algorithm. Otherwise, if treatment and control are factors then the optimized linear complexity algorithm is used, otherwise the RLE algorithm (with complexity n log n) is used.

Value

A list of class effsize containing the following components:

- estimate: the Cliff’s delta estimate
- conf.int: the confidence interval of the delta
- var: the estimated variance of the delta
- conf.level: the confidence level used to compute the confidence interval
- dm: the dominance matrix used for computation, only if return.dm is TRUE
- magnitude: a qualitative assessment of the magnitude of effect size
- method: the method used for computing the effect size, always "Cliff’s Delta"
- variance.estimation: the method used to compute the delta variance estimation, either "unbiased" or "consistent"
- CI.distribution: the distribution used to compute the confidence interval, either "Normal" or "Student-t"

The magnitude is assessed using the thresholds provided in (Romano 2006), i.e. |d|<0.147 “negligible”, |d|<0.33 ”small”, |d|<0.474 ”medium”, otherwise ”large”

Author(s)

Marco Torchiano http://softeng.polito.it/torchiano/

References


See Also

cohen.d, print.effsize
cohen.d

Examples

```r
## Example data from Hogarty and Kromrey (1999)
treatment <- c(10, 10, 20, 20, 30, 30, 40, 50)
control <- c(10, 20, 30, 40, 50)
res = cliff.delta(treatment, control, return.dm=TRUE)
print(res)
print(res$dm)
```  

---

cohen.d  

*Cohen’s d and Hedges’ g effect size*

Description

Computes the Cohen’s d and Hedges’ g effect size statistics.

Usage

```r
cohen.d(d, ...)
```

```r
## S3 method for class 'formula'
cohen.d(formula, data=list(),...)
```

```r
## Default S3 method:
cohen.d(d, f, pooled=TRUE, paired=FALSE,
       na.rm=FALSE, hedges.correction=FALSE,
       conf.level=0.95, ...)
```

Arguments

- **d**: a numeric vector giving either the data values (if `f` is a factor) or the treatment group values (if `f` is a numeric vector)
- **f**: either a factor with two levels or a numeric vector of values
- **pooled**: a logical indicating whether compute pooled standard deviation or the whole sample standard deviation
- **paired**: *deprecated* a logical indicating whether to consider the values as paired  
  *Since version 0.3.1 this parameter is ignored and may be removed in future code clean-ups*
- **na.rm**: logical indicating whether NA should be removed before computation
- **hedges.correction**: logical indicating whether apply the Hedges correction
- **conf.level**: confidence level of the confidence interval
- **formula**: a formula of the form `y ~ f`, where `y` is a numeric variable giving the data values and `f` a factor with two levels giving the corresponding groups
data: an optional matrix or data frame containing the variables in the formula. By default the variables are taken from environment(formula).

... further arguments to be passed to or from methods.

Details

When f in the default version is a factor or a character, it must have two values and it identifies the two groups to be compared. Otherwise (e.g. f is numeric), it is considered as a sample to be compare to d.

In the formula version, if f is expected to be a factor, if that is not the case it is coherced to a factor and a warning is issued.

The function computes the value of Cohen’s d statistics (Cohen 1988). If required (hedges.correction==TRUE) the Hedges g statistics is computed instead (Hedges and Holkin, 1985).

Also a quantification of the effect size magnitude is performed using the thresholds define in Cohen (1992). The magnitude is assessed using the thresholds provided in (Cohen 1992), i.e. ld1<0.2 "negligible", ld1<0.5 "small", ld1<0.8 "medium", otherwise "large"

The variance of the d is computed using the conversion formula reportead at page 238 of Cooper et al. (2009):

\[
S_d^2 = \left( \frac{n_1 + n_2}{n_1 n_2} + \frac{d^2}{2df} \right) \left( \frac{n_1 + n_2}{df} \right)
\]

Value

A list of class effsize containing the following components:

- estimate: the statistics estimate
- conf.int: the confidence interval of the statistic
- var: the estimated variance of the statistic
- conf.level: the confidence level used to compute the confidence interval
- magnitude: a qualitative assessment of the magnitude of effect size
- method: the method used for computing the effect size, either "Cohen’s d" or "Hedges’ g"

Author(s)

Marco Torchiano http://softeng.polito.it/torchiano/

References


The Handbook of Research Synthesis and Meta-Analysis (Cooper, Hedges, & Valentine, 2009)
See Also

\texttt{cliff.delta}, \texttt{VD.A}, \texttt{print.effsize}

Examples

treatment = rnorm(100,mean=10)
control = rnorm(100,mean=12)
d = c(treatment,control))
f = rep(c("Treatment","Control"),each=100)
## compute Cohen's d
## treatment and control
cohen.d(treatment,control)
## data and factor
cohen.d(d,f)
## formula interface
cohen.d(d ~ f)
## compute Hedges' g
cohen.d(d,f,hedges.correction=TRUE)

---

\texttt{print.effsize} \hspace{1cm} \textit{Prints effect size}

Description

Prints the results of an effect size computation

Usage

\texttt{## S3 method for class 'effsize'
print(x, ...)}

Arguments

\texttt{x} \hspace{1cm} the effect size result
\texttt{...} \hspace{1cm} further parameters are currently ignored

Details

Shows the estimate value and, when available, the confidence interval.

Note

This is still work in progress..

Author(s)

Marco Torchiano \url{http://softeng.polito.it/torchiano/}
References

See the main function \texttt{cliff.delta}.

See Also

\texttt{cliff.delta} \texttt{cohen.d}

---

\textbf{VD.A} \hfill \textit{Vargha and Delaney A measure}

Description

Computes the Vargha and Delaney A effect size measure.

Usage

\begin{verbatim}
VD.A(d, ...)  

## S3 method for class 'formula'
VD.A(formula, data=list(), ...)

## Default S3 method:
VD.A(d, f, ...)
\end{verbatim}

Arguments

- \texttt{d}: a numeric vector giving either the data values (if \texttt{f} is a factor) or the treatment group values (if \texttt{f} is a numeric vector)
- \texttt{f}: either a factor with two levels or a numeric vector of values
- \texttt{formula}: a formula of the form \texttt{y \sim f}, where \texttt{y} is a numeric variable giving the data values and \texttt{f} a factor with two levels giving the corresponding group
- \texttt{data}: an optional matrix or data frame containing the variables in the formula \texttt{formula}. By default the variables are taken from \texttt{environment(formula)}.
- \texttt{...}: further arguments to be passed to or from methods.

Details

The function computes the Vargha and Delaney A effect size measure (Vargha and Delaney, 2000).

Value

A list of class \texttt{effsize} containing the following components:

- \texttt{estimate}: the A statistics estimate
- \texttt{magnitude}: a qualitative assessment of the magnitude of effect size
- \texttt{method}: the method used, i.e. "Vargha and Delaney A"
Author(s)
Marco Torchiano http://softeng.polito.it/torchiano/

References

See Also
cliff.delta, cohen.d, print.effsize

Examples
```r
treatment = rnorm(100, mean=10)
control = rnorm(100, mean=12)
d = c(treatment, control)
f = rep(c("treatment", "control"), each=100)
## compute Vargha and Delaney A
## treatment and control
VD.A(treatment, control)
## data and factor
VD.A(d, f)
## formula interface
VD.A(d ~ f)
```
Index

*Topic **Cliff**
  cliff.delta, 3

*Topic **Cohen**
  cohen.d, 5

*Topic **Hedges**
  cohen.d, 5

*Topic **Vargha and Delaney**
  VD.A, 8

*Topic **effect size**
  cliff.delta, 3
  cohen.d, 5
  effsize-package, 2
  VD.A, 8

*Topic **package**
  effsize-package, 2
  cliff.delta, 2, 3, 7–9
  cohen.d, 2, 4, 5, 8, 9
  effsize (effsize-package), 2
  effsize-package, 2
  print.effsize, 4, 7, 7, 9
  VD.A, 2, 7, 8