

# Package ‘tatest’

July 18, 2018

**Type** Package

**Title** Two-Group Ta-Test

**Version** 1.0

**Date** 2018-07-10

**Author** Yuan-De Tan

**Maintainer** Yuan-De Tan <tanyuande@gmail.com>

**Description** The ta-test is a modified two-sample or two-group t-test of Gosset (1908). In small samples with less than 15 replicates, the ta-test significantly reduces type I error rate but has almost the same power with the t-test and hence can greatly enhance reliability or reproducibility of discoveries in biology and medicine. The ta-test can test single null hypothesis or multiple null hypotheses without needing to correct p-values.

**License** GPL-3

**Depends** R (>= 3.3.0), stats, base, utils

**LazyLoad** yes

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2018-07-18 11:30:05 UTC

## R topics documented:

tatest-package . . . . .	2
Ecadherin . . . . .	2
momega . . . . .	3
omega . . . . .	4
rhov . . . . .	6
ta.test . . . . .	7
ttest . . . . .	9

<b>Index</b>	<b>12</b>
--------------	-----------

---

tatest-package            *Two-Group  $T_{\alpha}$ -Test*

---

**Description**

The  $t_{\alpha}$ -test is a modified two-sample or two-group t-test of Gosset(1908).

**Details**

The package tatest consists of three t-tests: ttest, tatest and mtatest. ttest, rhov,and omega construct tatest, ttest and rhov and momega construct mtatest.  $t_{\alpha}$ -test is used to test for a single null hypothesis and  $mt_{\alpha}$ -test is used to identify differential expressed genes in microarray data and RPPA data or proteomic data.

**Author(s)**

Yuande Tan  
<tanyuande@gmail.com> Yuande Tan  
Maintainer: <tanyuande@gmail.com>

**See Also**

[t.test](#)

**Examples**

```
X<-c(112,122,108,127)
Y<-c(302, 314,322,328)
dat<-cbind(X,Y)
ta.test(X=dat,na=4,nb=4, eqv=TRUE, paired=FALSE,
alpha=0.05, alternative = "two.sided",LOG="NULL")
t.test(x=X,y=Y)
ttest(X,Y,alternative = "two.sided", LG = "NULL")
```

---

Ecadherin            *A microarray data*

---

**Description**

Analysis of the effects of loss of E-cadherin and cell adhesion on human mammary epithelial cells

**Usage**

```
data("Ecadherin")
```

**Format**

A data frame with 5000 observations on the following 7 variables.

geneid a factor with levels  
shCntrl\_rep1 a numeric vector  
shCntrl\_rep2 a numeric vector  
shCntrl\_rep3 a numeric vector  
shEcad\_rep1 a numeric vector  
shEcad\_rep2 a numeric vector  
shEcad\_rep3 a numeric vector

**Details**

Loss of the epithelial adhesion molecule E-cadherin is thought to enable metastasis by disrupting intercellular contacts—an early step in metastatic dissemination. To further investigate the molecular basis of this notion, two methods were used to inhibit E-cadherin function that distinguish between E-cadherin's cell-cell adhesion and intracellular signaling functions. While the disruption of cell-cell contacts alone does not enable metastasis, the loss of E-cadherin protein does, through induction of an epithelial-to-mesenchymal transition, invasiveness and anoikis-resistance. In addition, gene expression analysis shows that E-cadherin loss results in the induction of multiple transcription factors, at least one of which, Twist, is necessary for E-cadherin loss-induced metastasis. These findings indicate that E-cadherin loss in tumors contributes to metastatic dissemination by inducing wide-ranging transcriptional and functional changes. This dataset consists 5000 of 22277 gene libraries, two conditions shCntrl and shEcad, each having 3 replicates. With access number GSE9691, you can obtain a whole data of 22277 gene libraries in GEO.

**Examples**

```
data(Ecadherin)
```

---

momega

*Null rho for testing multiple null hypotheses*

---

**Description**

Momega is a function that is used to estimate omega for testing multiple null hypotheses. Omega is a null  $\rho$  that is used as a threshold for  $\rho$  from real data. Since this omega is majorly applied to microarray, we named this function as momega.

**Usage**

```
momega(dat, na, nb, distr = "norm")
```

**Arguments**

dat	a matrix dataset containing string columns and two-condition numeric columns
na	numeric value, size of sample 1 or number of replicates in condition A.
nb	numeric value, size of sample 1 or number of replicates in condition A.
distr	string, data distribution.

**Details**

This function is to use null data to calculate omega value with  $\rho = 1$ .

**Value**

return a numeric value

**Author(s)**

Yuan-De Tan <tanyuande@gmail.com>

**References**

Yuan-De Tan Anita M. Chandler, Arindam Chaudhury, and Joel R. Neilson (2015) A Powerful Statistical Approach for Large-scale Differential Transcription Analysis. *Plos One*. 2015 DOI: 10.1371/journal.pone.0123658.

**See Also**

[omega,rhov](#)

**Examples**

```
data(Ecadherin)
Ecad<-as.data.frame(Ecadherin)
Ecad1<-Ecad[,-1]
w<-momega(dat=Ecad1[1:2000,], na=3, nb=3, distr = "norm")
```

---

omega

*Omega calcularion*

---

**Description**

Omega ( $\omega$ ) function is a function that is used to estimate omega using simulate null data from negative binomial distribution. Omega is a null rho that is used as a threshold for real rho. Simulation is dependent on the original data.

**Usage**

```
omega(XA, XB, na, nb, m, alpha = 0.05, distr = "norm")
```

**Arguments**

XA	a numeric vector values of condition A
XB	a numeric vector values of condition b
na	a numeric value, sample size in A.
nb	a numeric value, sample size in B
m	simulation number specified.
alpha	a numeric value, statistical cutoff. Default is 0.05.
distr	data distribution specified for data. For current version, we consider three distributions: normal, negative binomial (NB) and uniform. Default distribution is "norm". If user believe that the data follow negative binomial distribution, then set distr="NB" or "negative binomial" or if the data follow uniform distribution, then set distr="unif" or "uniform".

**Details**

This function is to use simulated null data to calculate omega value with  $\rho = 1$ .

**Value**

return a numeric value

**Author(s)**

Yuan-De Tan <tanyuande@gmail.com>

**References**

Yuan-De Tan Anita M. Chandler, Arindam Chaudhury, and Joel R. Neilson(2015) A Powerful Statistical Approach for Large-scale Differential Transcription Analysis. *Plos One*. 2015 DOI: 10.1371/journal.pone.0123658.

**See Also**

[momega](#), [rhov](#)

**Examples**

```
X<-c(112,122,108,127)
Y<-c(302, 314,322,328)
omega(XA=X, XB=Y, na=4, nb=4, m=2000, alpha = 0.05)
#[1] 0.9055152
omega(XA=X, XB=Y, na=4, nb=4, m=2000, alpha = 0.05,distr="NB")
#[1] 0.8995424
omega(XA=X, XB=Y, na=4, nb=4, m=2000, alpha = 0.05,distr="uniform")
#[1] 0.97194
```



**Examples**

```
X<-c(112,122, 108,127)
Y<-c(302, 314, 322, 328)
rhov(XA=X, XB=Y)
#[1] 3.062173
```

---

ta.test	<i>t<sub>α</sub>-test for testing difference between two conditions with small samples</i>
---------	--

---

**Description**

$t_{\alpha}$ -test is a modified t-test of Gosset(1908).

**Usage**

```
ta.test(X, nci=NULL, na, nb, alpha=0.05, paired=FALSE, eqv=TRUE,
LOG=c("NULL", "LOG2", "LOG", "LOG10"),
alternative = c("two.sided", "less", "greater"), distr="norm")
```

**Arguments**

X	a numeric vector data or a numeric matrix (two columns) dataset with two conditions or groups A and B having sample size na and nb, respectively. X is also a datasheet of multiple variables (for example, genes ) with two groups A and B and information columns(see an example Ecadherin)
nci	a numeric int value, column numbers for strings or information of data. If X is a numeric vector data or a numeric matrix (two columns) dataset, then nci=NULL, otherwise, nci is non-zero int, meaning that at least one column is reserved as information of data
na	numeric constant, sample size in condition A
nb	numeric constant, sample size in condition B
alpha	a numeric constant, statistical cutoff for significance level. If dataset X is multiple variables (rows) (variable number is over 1000), then alpha is taken default values of 0.05 and 0.01.
paired	a logical indicating whether you want a paired t-test.
eqv	a logical variable indicating whether to treat the two variances as being equal. If TRUE then the pooled variance is used to estimate the variance otherwise the Welch (or Satterthwaite) approximation to the degrees of freedom is used.
LOG	logarithm: if data are not taken logarithm conversion, then LOG=NULL , if data are converted into logarithm with base 10, 2 or natural logarithm, then correspondingly, LOG="LOG10", "LOG2" or "LOG".
alternative	a character string specifying the alternative hypothesis, must be one of "two.sided" (default), "greater" or "less". User can specify just the initial letter.

**distr** data distribution for estimating omega ( $\omega$ ). Default is normal distribution. In current version, we just consider three types of distributions: "negative distribution"(or "NB"), "norm"(or "normal"), and "unif"("uniform"). User can choose one of them according to user's data distribution.

### Details

The formula interface is only applicable for the 2-sample tests.

**alternative** = "greater" is the alternative that x has a larger mean than y.

If **paired** is TRUE then both x and y must be specified and they must be the same length. Missing values are silently removed (in pairs if **paired** is TRUE).

If **var.equal** is TRUE then the pooled estimate of the variance is used. By default, if **var.equal** is FALSE, then the variance is estimated separately for both groups and the Welch modification to the degrees of freedom is used.

If the input data are effectively constant (compared to the larger of the two means) an error is generated.

When pooled variance is larger, **ttest** is equivalent to **t.test**; When pooled variance is small, **t-tvalue** of **ttest** is smaller than that of **t.test**; When the pooled variance is zero, **t-value** of **t.test** is not defined but **tvalue** of **ttest** is still defined.

**Omega** is dependent on **alpha** value.

**ta.test** can be used to test for differentially expressed genes in microarray data, proteomic data and RPPA data but cannot be used to count data of RNA-seq reads.

### Value

If X is single-hypothesis test data, then a list similar to **t.test** containing the following components:

**statistic**

the value of the t-statistic.

**parameter**

the degrees of freedom for the t-statistic.

**p.value**

the p-value for the test.

**conf.int**

a confidence interval for the mean appropriate to the specified alternative hypothesis.

**estimate**

the estimated mean or difference in means depending on whether it was a one-sample test or a two-sample test.

**null.value**

the specified hypothesized value of the mean or mean difference depending on whether it was a one-sample test or a two-sample test.

**alternative**

a character string describing the alternative hypothesis.

**method**

a character string indicating what type of t-test was performed.

**data.name**

a character string giving the name(s) of the data.

**omega** a threshold for rho in **tatest**.

If X is microarray, RPPA data, then a list is **datasheet** including **t01**, **pv01**, **t05** and **p05** columns



**Note**

LOG choice must be the same with log transformation of data.

**Author(s)**

Yuande Tan  
<tanyuande@gmail.com>

**References**

Gosset, W.S, "Probable error of a correlation coefficient". Biometrika. 1908,6 (2/3): 302-310.

**See Also**

[t.test](#), [ttest](#)

**Examples**

```
X<-c(112,122,108,127)
Y<-c(302, 314,322,328)
dat<-cbind(X,Y)
ta.test(X=dat,nci=NULL, na=4,nb=4, eqv=TRUE,
alpha=0.05, LOG="NULL", alternative = "two.sided")

data(Ecadherin)
res<-ta.test(X=Ecadherin[1:2000,],nci=1, na=3,nb=3, eqv=TRUE,
LOG="LOG2", alternative = "two.sided")
```

---

ttest

*A modified t-test*


---

**Description**

ttest is a modified t-test of Gosset(1098).

**Usage**

```
ttest(x, y = NULL, alternative = c("two.sided", "less", "greater"),
mu = 0, paired = FALSE, var.equal = FALSE, conf.level = 0.95,
LG = c("NULL", "LOG2", "LOG", "LOG10"), ...)
```

**Arguments**

x a (non-empty) numeric vector of data values.  
y an optional (non-empty) numeric vector of data values. For  $t_{\alpha}$ -test, y is required  
alternative a character string specifying the alternative hypothesis, must be one of "two.sided" (default), "greater" or "less". You can specify just the initial letter.

<code>mu</code>	a number indicating the true value of the mean (or difference in means if you are performing a two sample test).
<code>paired</code>	a logical indicating whether you want a paired t-test.
<code>var.equal</code>	a logical variable indicating whether to treat the two variances as being equal. If TRUE then the pooled variance is used to estimate the variance otherwise the Welch (or Satterthwaite) approximation to the degrees of freedom is used.
<code>conf.level</code>	confidence level of the interval.
<code>LG</code>	logarithm: if x and y are not taken logarithm conversion, the LG=NULL, if x and y are converted into logarithm with base 10, 2 or natural logarithm, then correspondingly, LG="LOG10",
<code>...</code>	further arguments to be passed to or from methods.

### Details

The formula interface is only applicable for the 2-sample tests.

`alternative = "greater"` is the alternative that x has a larger mean than y.

If `paired` is TRUE then both x and y must be specified and they must be the same length. Missing values are silently removed (in pairs if `paired` is TRUE).

If `var.equal` is TRUE then the pooled estimate of the variance is used. By default, if `var.equal` is FALSE, then the variance is estimated separately for both groups and the Welch modification to the degrees of freedom is used.

If the input data are effectively constant (compared to the larger of the two means) an error is generated.

When pooled variance is larger, `ttest` is equivalent to `t.test`; When pooled variance is small, t-tvalue of `ttest` is smaller than that of `t.test`; When the pooled variance is zero, t-value of `t.test` is not defined but tvalue of `ttest` is still defined.

### Value

A list with class "htest" containing the following components:

`statistic`

the value of the t-statistic.

`parameter`

the degrees of freedom for the t-statistic.

`p.value`

the p-value for the test.

`conf.int`

a confidence interval for the mean appropriate to the specified alternative hypothesis.

`estimate`

the estimated mean or difference in means depending on whether it was a one-sample test or a two-sample test.

`null.value`

the specified hypothesized value of the mean or mean difference depending on whether it was a one-sample test or a two-sample test.

`alternative`

a character string describing the alternative hypothesis.

`method`

a character string indicating what type of t-test was performed.  
data.name a character string giving the name(s) of the data.

**Author(s)**

Yuande Tan  
<tanyuande@gmail.com>

**References**

Gosset, W.S, "Probable error of a correlation coefficient". Biometrika. 1908,6(2/3): 302-310.

**See Also**

[stats{t.test}](#), [ta.test](#)

**Examples**

```
# The following example is log10 transformed data
YA<-c(4.1455383,4.7186418,4.6483316)
YB<-c(6.9184465,5.609667,6.5920944)
t.test(x=YA, y=YB)
ttest(YA,YB,alternative = "two.sided", LG = "NULL")
ttest(YA,YB,alternative = "two.sided", LG = "LOG2")
ttest(YA,YB,alternative = "two.sided", LG = "LOG10")
```

# Index

- \*Topic **dataset**
    - Ecadherin, [2](#)
  - \*Topic **gap**
    - rhov, [6](#)
  - \*Topic **multiple tests**
    - momega, [3](#)
  - \*Topic **noise**
    - rhov, [6](#)
  - \*Topic **null rho**
    - momega, [3](#)
  - \*Topic **rhov**
    - ta.test, [7](#)
  - \*Topic **rho**
    - omega, [4](#)
  - \*Topic **t.test**
    - ttest, [9](#)
  - \*Topic **tatest**
    - tatest-package, [2](#)
  - \*Topic **thresholds**
    - omega, [4](#)
  - \*Topic **ttest**
    - ta.test, [7](#)
    - ttest, [9](#)
- Ecadherin, [2](#)
- momega, [3](#), [5](#)
- omega, [4](#), [4](#), [6](#)
- rhov, [4](#), [5](#), [6](#)
- stats, [11](#)
- t.test, [2](#), [9](#)
- ta.test, [7](#), [11](#)
- tatest-package, [2](#)
- ttest, [9](#), [9](#)