Package: tcftt (via r-universe)

September 1, 2024

Type Package

Title Two-Sample Tests for Skewed Data

Version 0.1.0

Author Huaiyu Zhang, Haiyan Wang

Maintainer Huaiyu Zhang <huaiyuzhang1988@gmail.com>

Description The classical two-sample t-test works well for the normally distributed data or data with large sample size. The tcfu() and tt() tests implemented in this package provide better type-I-error control with more accurate power when testing the equality of two-sample means for skewed populations having unequal variances. These tests are especially useful when the sample sizes are moderate. The tcfu() uses the Cornish-Fisher expansion to achieve a better approximation to the true percentiles. The tt() provides transformations of the Welch's t-statistic so that the sampling distribution become more symmetric. For more technical details, please refer to Zhang (2019) <http://hdl.handle.net/2097/40235>.

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 7.1.0

Depends R (>= 3.1.0)

Imports stats

Repository https://huaiyuzhang.r-universe.dev

RemoteUrl https://github.com/huaiyuzhang/tcftt

RemoteRef HEAD

RemoteSha e12bfd9ed347fc2d8c61ae27ef565ba05d05ecec

Contents

---

adjust_power   *Adjusting power to assure actual size is within significance level*

Description

It is common to use Monte Carlo experiments to evaluate the performance of hypothesis tests and
compare the empirical power among competing tests. High power is desirable but difficulty arises
when the actual sizes of competing tests are not comparable. A possible way of tackling this issue
is to adjust the empirical power according to the actual size. This function incorporates three types
of power adjustment methods.

Usage

```
adjust_power(size, power, method = "ZW")
```

Arguments

size	the empirical size of a test.
power	the empirical power of a test.
method	the power adjustment method. 'ZW' is the method proposed by Zhang and Wang (2020), 'CYS' is the method proposed by Cavus et al. (2019), and 'probit' is the "method 1: probit analysis" in Lloyd (2005).

Value

the power value after adjustment.

References

Lloyd, C. J. (2005). Estimating test power adjusted for size. Journal of Statistical Computation and
Simulation, 75(11):921-933.

Cavus, M., Yazici, B., & Sezer, A. (2019). Penalized power approach to compare the power of the
tests when Type I error probabilities are different. Communications in Statistics-Simulation and
Computation, 1-15.

Zhang, H. and Wang, H. (2020). Transformation tests and their asymptotic power in two-sample
comparisons Manuscript in review.

Examples

```
adjust_power(size = 0.06, power = 0.8, method = 'ZW')
adjust_power(size = 0.06, power = 0.8, method = 'CYS')
adjust_power(size = 0.06, power = 0.8, method = 'probit')
```

---

boot_test   *Bootstrap_t test for two-sample comparisons*

Description

This function provides bootstrap approximation to the sampling distribution of the the Welch's t-statistic

Usage

```
boot_test(x1, x2, B = 1000, alternative = "greater")
```

Arguments

x1	the first sample.
x2	the second sample.
B	number of resampling rounds. Default value is 1000.
alternative	the alternative hypothesis: "greater" for upper-tailed, "less" for lower-tailed, and "two.sided" for two-sided alternative.

Value

the p-value of the bootstrap_t test.

Examples

```
x1 <- rnorm(100, 0, 1)
x2 <- rnorm(100, 0.5, 2)
boot_test(x1, x2)
```

---

pauc   *Power-adjustment based on non-parametric estimation of the ROC curve*

Description

It is common to use Monte Carlo experiments to evaluate the performance of hypothesis tests and compare the empirical power among competing tests. High power is desirable but difficulty arises when the actual sizes of competing tests are not comparable. A possible way of tackling this issue is to adjust the empirical power according to the actual size. This function implements the "method 2: non-parametric estimation of the ROC curve" in Lloyd (2005). For more details, please refer to the paper.

Usage

```
pauc(stat_h0, stat_ha, target_range_lower, target_range_upper)
```

Arguments

stat_h0   simulated test statistics under the null hypothesis.

stat_ha   simulated test statistics under the alternative hypothesis.

target_range_lower
                    the lower end of the size range.

target_range_upper
                    the upper end of the size range.

Value

the adjusted power.

References

Lloyd, C. J. (2005). Estimating test power adjusted for size. Journal of Statistical Computation and Simulation, 75(11):921-933.

Examples

```
stath0 <- rnorm(100)
statha <- rnorm(100, mean=1)
pauc(stath0, statha, 0.01, 0.1)
```

---

tcftt *tcftt: Two-Sample Tests for Skewed Data*

Description

The classical two-sample t-test works well for the normally distributed data or data with large sample size. The tcfu() and tt() tests implemented in this package provide better type I error control with more accurate power when testing the equality of two-sample means for skewed populations having unequal variances. The approximation is especially useful when the sample sizes are moderate. The tcfu() uses the Cornish-Fisher expansion to achieve a better approximation to the true percentiles. The tt() provides transformations of the Welch's t-statistic so that the sampling distribution become more symmetric. For more technical details, please refer to Zhang (2019) <http://hdl.handle.net/2097/40235>.

tcftt functions

The function 'tcfu()' implements the Cornish-Fisher based two-sample test (TCFU) and 'tt()' implements the transformation based two-sample test (TT). The function 't_edgeworth()' provides the Edgeworth expansion for cumulative distribution function for the Welch's t-statistic, and 't_cornish_fisher()' provides the Cornish-Fisher expansion for the percentiles. The functions 'adjust_power()' and 'pauc()' provide power adjustment for simulation studies so that the actual size of the tests are within the significance level.

---

tcfu *The TCFU test*

Description

This test is suitable for testing the equality of two-sample means for the populations having unequal variances. When the populations are not normally distributed, this test can provide better type I error control and more accurate power than a large-sample t-test using normal approximation. The critical values of the test are computed based on the Cornish-Fisher expansion of the Welch's t-statistic. The order of the Cornish-Fisher expansion is allowed to be 0, 1, or 2. More details please refer to Zhang and Wang (2020).

Usage

```
tcfu(x1, x2, effectSize = 0, alternative = "greater", alpha = 0.05, order = 2)
```

Arguments

x1	the first sample.
x2	the second sample.
effectSize	the effect size of the test. The default value is 0.

alternative	the alternative hypothesis: "greater" for upper-tailed, "less" for lower-tailed, and "two.sided" for two-sided alternative.
alpha	the significance level. The default value is 0.05.
order	the order of the Cornish-Fisher expansion.

Value

test statistic, critical value, p-value, reject decision at the given significance level.

References

Zhang, H. and Wang, H. (2020). Transformation tests and their asymptotic power in two-sample comparisons. Manuscript in review.

Examples

```
x1 <- rnorm(20, 1, 3)
x2 <- rnorm(21, 2, 3)
tcfu(x1, x2, alternative = 'two.sided')
```

---

tt	*The transformation based test*

Description

This test is suitable for testing the equality of two-sample means for the populations having unequal variances. When the populations are not normally distributed, the sampling distribution of the Welch's t-statistic may be skewed. This test conducts transformations of the Welch's t-statistic to make the sampling distribution more symmetric. For more details, please refer to Zhang and Wang (2020).

Usage

```
tt(x1, x2, alternative = "greater", effectSize = 0, alpha = 0.05, type = 1)
```

Arguments

x1	the first sample.
x2	the second sample.
alternative	the alternative hypothesis: "greater" for upper-tailed, "less" for lower-tailed, and "two.sided" for two-sided alternative.
effectSize	the effect size of the test. The default value is 0.
alpha	the significance level. The default value is 0.05.

type the type of transformation to be used. Possible choices are 1 to 4. They correspond to the TT1 to TT4 in Zhang and Wang (2020). Which type provides the best test depends on the relative skewness parameter A in Theorem 2.2 of Zhang and Wang (2020). In general, if A is greater than 3, type =3 is recommended. Otherwise, type=1 or 4 is recommended. The type=2 transformation may be more conservative in some cases and more liberal in some other cases than the type=1 and 4 transformations. For more details, please refer to Zhang and Wang (2020).

Value

test statistic, critical value, p-value, reject decision at the given significance level.

References

Zhang, H. and Wang, H. (2020). Transformation tests and their asymptotic power in two-sample comparisons Manuscript in review.

Examples

```
x1 <- rnorm(20, 1, 3)
x2 <- rnorm(21, 2, 3)
tt(x1, x2, alternative = 'two.sided', type = 1)

#Negative lognormal versus normal data
 n1=50;  n2=33
 x1 = -rlnorm(n1, meanlog = 0, sdlog = sqrt(1)) -0.3*sqrt((exp(1)-1)*exp(1))
 x2 = rnorm(n2, -exp(1/2), 0.5)
 tt(x1, x2, alternative = 'less', type = 1)
 tt(x1, x2, alternative = 'less', type = 2)
 tt(x1, x2, alternative = 'less', type = 3)
 tt(x1, x2, alternative = 'less', type = 4)

#Lognormal versus normal data
 n1=50;  n2=33
 x1 = rlnorm(n1, meanlog = 0, sdlog = sqrt(1)) + 0.3*sqrt((exp(1)-1)*exp(1))
 x2 = rnorm(n2, exp(1/2), 0.5)
 tt(x1, x2, alternative = 'greater', type = 1)
 tt(x1, x2, alternative = 'greater', type = 2)
 tt(x1, x2, alternative = 'greater', type = 3)
 tt(x1, x2, alternative = 'greater', type = 4)
```

---

t_cornish_fisher *Cornish-Fisher expansion for Welch's t-statistic*

Description

This function provides approximation for the quantile function of the sampling distribution of the Welch's t-statistic using Cornish-Fisher expansion (up to second order).

Usage

```
t_cornish_fisher(
  p,
  order = 2,
  n1,
  n2,
  mu1,
  mu2,
  sigma1,
  sigma2,
  gamma1,
  gamma2,
  tau1,
  tau2
)
```

Arguments

p	a probability value.
order	the order of Cornish-Fisher expansion. Valid options are 0, 1, and 2. If set to 0, it reduces to a normal approximation and it returns the p-th percentile of standard normal distribution.
n1	sample size for the sample from the first population.
n2	sample size for the sample from the second population.
mu1	mean of the first population.
mu2	mean of the second population.
sigma1	standard deviation of the first population.
sigma2	standard deviation of the second population.
gamma1	skewness of the first population.
gamma2	skewness of the second population.
tau1	kurtosis of the first population.
tau2	kurtosis of the second population.

Value

Cornish-Fisher expansion value evaluated at p.

Examples

```
t_cornish_fisher(0.9, order=2,
n1=60, n2=30,
mu1=0, mu2=0,
sigma1=1, sigma2=0.5,
gamma1=1, gamma2=0,
tau1=6, tau2=0)
```

```
t_cornish_fisher(0.3, order=1,
n1=60, n2=30,
mu1=0, mu2=0,
sigma1=1, sigma2=0.5,
gamma1=1, gamma2=0,
tau1=6, tau2=0)
```

---

t_edgeworth   *Edgeworth expansion for Welch's t-statistic*

Description

This function provides approximation for the cumulative distribution function of the sampling distribution of the Welch's t-statistic using Normal distribution, first order or second order Edgeworth expansion.

Usage

```
t_edgeworth(
  x,
  order = 2,
  n1,
  n2,
  mu1,
  mu2,
  sigma1,
  sigma2,
  gamma1,
  gamma2,
  tau1,
  tau2
)
```

Arguments

x	a real number.
order	the order of edgeworth expansion. Valid options are 0, 1, and 2. If set to 0, it reduces to approximation based on the central limit theorem and returns the CDF of standard normal distribution evaluated at x.
n1	sample size for the sample from the first population.
n2	sample size for the sample from the second population.
mu1	mean of the first population.
mu2	mean of the second population.
sigma1	standard deviation of the first population.
sigma2	standard deviation of the second population.

gamma1	skewness of the first population.
gamma2	skewness of the second population.
tau1	kurtosis of the first population.
tau2	kurtosis of the second population.

Value

Edgeworth expansion evaluated at x.

Examples

```
t_edgeworth(1.96, order=2,
n1=20, n2=30,
mu1=0, mu2=0,
sigma1=1, sigma2=0.5,
gamma1=1, gamma2=0,
tau1=6, tau2=0)
```

Index