Package: wmwm (via r-universe)

September 7, 2024

Title Performs Wilcoxon-Mann-Whitney Test with Missing Data

Version 1.0.0

Description Performs Wilcoxon-Mann-Whitney test in the presence of missing data with controlled Type I error regardless of the values of missing data.

License MIT + file LICENSE

Depends R (>= 3.2.1)

Imports stats (>= 3.2.1)

Suggests spelling, testthat (>= 3.0.0)

Config/testthat/edition 3

Encoding UTF-8

RoxygenNote 7.3.1

Language en-US

NeedsCompilation no

Author Yijin Zeng [aut, cre, cph], Dean Bodenham [aut], Niall Adams [aut]

Maintainer Yijin Zeng <`yijinzeng98@gmail.com`>

Date/Publication 2024-07-27 16:10:02 UTC

Repository https://yijin-zeng.r-universe.dev

RemoteUrl https://github.com/cran/wmwm

RemoteRef HEAD

RemoteSha c49edc6d22c315afb06ea79f0239f9506733dc56

Contents

`wmwm.test` *Wilcoxon-Mann-Whitney Test in the Presence of Arbitrarily Missing Data*

Description

Performs the two-sample Wilcoxon-Mann-Whitney test in the presence of missing data, which controls the Type I error regardless of the values of missing data.

Usage

```
wmwm.test(X, Y, alternative = c("two.sided", "less", "greater"),
ties = NULL, lower.boundary = -Inf, upper.boundary = Inf,
exact = NULL, correct = TRUE)
```

Arguments

`X, Y`	numeric vectors of data values with potential missing data. Inf and -Inf values will be omitted.
`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.
`ties`	a logical indicating whether samples could be tied. • If observed samples contain tied samples, ties defaults to TRUE. • If observed samples do not contain tied samples, ties defaults to FALSE.
`lower.boundary`	(when ties is TRUE) a number specifying the lower bound of the data set, must be smaller or equal than the minimum of all observed data.
`upper.boundary`	(when ties is TRUE) a number specifying the upper bound of the data set, must be larger or equal than the maximum of all observed data.
`exact`	a logical indicating whether the bounds should be of an exact p-value.
`correct`	a logical indicating whether the bounds should be of a p-value applying continuity correction in the normal approximation.

Details

`wmwm.test()` performs the two-sample hypothesis test method proposed in (Zeng et al., 2024) for univariate data when not all data are observed. Bounds of the Wilcoxon-Mann-Whitney test statistic and its p-value will be computed in the presence of missing data. The p-value of the test method proposed in (Zeng et al., 2024) is then returned as the maximum possible p-value of the Wilcoxon-Mann-Whitney test.

By default (if `exact` is not specified), this function returns bounds of an exact p-value if the length of `X` and `Y` are both smaller than 50, and there are no tied observations. Otherwise, bounds of a p-value calculated using normal approximation with continuity correction will be returned.

Value

`p.value`	the p-value for the test.
`bounds.statistic`	bounds of the value of the Wilcoxon-Mann-Whitney test statistic.
`bounds.pvalue`	bounds of the p-value of the Wilcoxon-Mann-Whitney test.
`alternative`	a character string describing the alternative hypothesis.
`ties.method`	a character string describing whether samples are considered tied.
`description.bounds`	a character string describing the bounds of the p-value.
`data.name`	a character string giving the names of the data.

References

- Zeng Y, Adams NM, Bodenham DA. On two-sample testing for data with arbitrarily missing values. arXiv preprint arXiv:2403.15327. 2024 Mar 22.
- Mann, Henry B., and Donald R. Whitney. "On a test of whether one of two random variables is stochastically larger than the other." The Annals of Mathematical Statistics (1947): 50-60.
- Lehmann, Erich Leo, and Howard J. D'Abrera. Nonparametrics: statistical methods based on ranks. Holden-day, 1975.

See Also

`stats::wilcox.test()` when data are fully observed.

Examples

```
#### Assume all samples are distinct.
X <- c(6.2, 3.5, NA, 7.6, 9.2)
Y <- c(0.2, 1.3, -0.5, -1.7)

## By default, when the sample sizes of both X and Y are smaller than 50,
## exact distribution will be used.
wmwm.test(X, Y, ties = FALSE, alternative = 'two.sided')

## using normality approximation with continuity correction:
wmwm.test(X, Y, ties = FALSE, alternative = 'two.sided', exact = FALSE, correct = TRUE)

#### Assume samples can be tied.
X <- c(6, 9, NA, 7, 9)
Y <- c(0, 1, 0, -1)

## When the samples can be tied, normality approximation will be used.
## By default, lower.boundary = -Inf, upper.boundary = Inf.
wmwm.test(X, Y, ties = TRUE, alternative = 'two.sided')

## specifying lower.boundary and upper.boundary:
wmwm.test(X, Y, ties = TRUE, alternative = 'two.sided', lower.boundary = -1, upper.boundary = 9)
```

Index