Package: wmwpow (via r-universe)

August 29, 2024

Type Package

Title Precise and Accurate Power of the Wilcoxon-Mann-Whitney Rank-Sum Test for a Continuous Variable

Version 0.1.3

Date 2020-07-19

Author Ilana Trumble, Orlando Ferrer, Camden Bay, Katie Mollan

Maintainer Ilana Trumble <ilana.trumble@cuanschutz.edu>

Description Power calculator for the two-sample Wilcoxon-Mann-Whitney rank-sum test for a continuous outcome (Mollan, Trumble, Reifeis et. al., Mar. 2020) <doi:10.1080/10543406.2020.1730866> <arXiv:1901.04597>, (Mann and Whitney 1947) <doi:10.1214/aoms/1177730491>, (Shieh, Jan, and Randles 2006) <doi:10.1080/10485250500473099>.

License GPL-3

Encoding UTF-8

LazyData true

Imports lamW, smoothmest, MASS

Depends R (>= 3.0.2)

RoxygenNote 7.1.1

NeedsCompilation no

Repository CRAN

Date/Publication 2020-07-20 23:40:02 UTC

Contents

| shiehpow | *Power Calculation Using the Shieh et. al. Approach* |

Description

The purpose of *shiehpow* is to perform a power analysis for a one or two-sided Wilcoxon-Mann-Whitney test using the method developed by Shieh and colleagues.

Arguments

n	Sample size of first sample (numeric)
m	Sample size of second sample (numeric)
p	Effect size, P(X<Y) (numeric)
alpha	Type I error rate (numeric)
dist	The distribution type for the two groups ("exp", "dexp", or "norm") (string)
sides	Options are "two.sided" and "one.sided" (string)

Note

When calculating power for dist="norm", *shiehpow* uses 100,000 draws from a $Z \sim N(0,1)$ distribution for the internal calculation of p2 and p3 from Shieh et al. (2006); thus *shiehpow* normal distribution power results may vary in the thousandths place from one run to the next.

References

Shieh, G., Jan, S. L., Randles, R. H. (2006). On power and sample size determinations for the Wilcoxon–Mann–Whitney test. Journal of Nonparametric Statistics, 18(1), 33-43.

Mollan K.R., Trumble I.M., Reifeis S.A., Ferrer O., Bay C.P., Baldoni P.L., Hudgens M.G. Exact Power of the Rank-Sum Test for a Continuous Variable, arXiv:1901.04597 [stat.ME], Jan. 2019.

Examples

```
# We want to calculate the statistical power to compare the distance between mutations on a DNA
# strand in two groups of people. Each group (X and Y) has 10 individuals. We assume that the
# distance between mutations in the first group is exponentially distributed with rate 3. We assume
# that the probability that the distance in the first group is less than the distance in the second
# group (i.e., P(X<Y)) is 0.8. The desired type I error is 0.05.

shiehpow(n = 10, m = 10, p = 0.80, alpha = 0.05, dist = "exp", sides = "two.sided")
```

---

wmwpowd	*Precise and Accurate Monte Carlo Power Calculation by Inputting Distributions F and G (wmwpowd)*

Description

wmwpowd has two purposes:

1. Calculate the power for a one-sided or two-sided Wilcoxon-Mann-Whitney test with an empirical p-value given two user specified distributions.

2. Calculate p, the P(X<Y), where X represents random draws from one continuous probability distribution and Y represents random draws from another distribution; p is useful for quantifying the effect size that the Wilcoxon-Mann-Whitney test is assessing.

Both 1. and 2. are calculated empirically using simulated data and output automatically.

Usage

```
wmwpowd(n, m, distn, distm, sides, alpha = 0.05, nsims = 10000)
```

Arguments

n	Sample size for the first distribution (numeric)
m	Sample size for the second distribution (numeric)
alpha	Type I error rate or significance level (numeric)
distn	Base R's name for the first distribution and any required parameters ("norm", "beta", "cauchy", "f", "gamma", "lnorm", "unif", "weibull","exp", "chisq", "t", "doublex")
distm	Base R's name for the second distribution and any required parameters ("norm", "beta", "cauchy", "f", "gamma", "lnorm", "unif", "weibull","exp", "chisq", "t", "doublex")
sides	Options are "two.sided", "less", or "greater". "less" means the alternative hypothesis is that distn is less than distm (string)
nsims	Number of simulated datasets for calculating power; 10,000 is the default. For exact power to the hundredths place (e.g., 0.90 or 90%) around 100,000 simulated datasets is recommended (numeric)

Note

Example of distn, distm: "norm(1,2)" or "exp(1)"

In addition to all continuous distributions supported in Base R, *wmwpowd* also supports the double exponential distribution from the smoothmest package

The output WMWOdds is p expressed as odds p/(1-p)

Use $ notation to select specific output parameters

The function has been optimized to run through simulations quickly; long wait times are unlikely for n and m of 50 or fewer

References

Mollan K.R., Trumble I.M., Reifeis S.A., Ferrer O., Bay C.P., Baldoni P.L., Hudgens M.G. Exact Power of the Rank-Sum Test for a Continuous Variable, arXiv:1901.04597 [stat.ME], Jan. 2019.

---

wmwpowp	*Precise and Accurate Monte Carlo Power Calculation by Inputting P (wmwpowp)*

Description

wmwpowp has two purposes:

1. Calculate the power for a one-sided or two-sided Wilcoxon-Mann-Whitney test with an empirical Monte Carlo p-value given one user specified distribution and p (defined as P(X<Y)).

2. Calculate the parameters of the second distribution. It is assumed that the second population is from the same type of continuous probability distribution as the first population.

Power is calculated empirically using simulated data and the parameters are calculated using derived mathematical formulas for P(X<Y).

Usage

```
wmwpowp(n, m, distn, k = 1, p = NA, wmwodds = NA, sides, alpha = 0.05, nsims = 10000)
```

Arguments

n	Sample size for the first distribution (numeric)
m	Sample size for the second distribution (numeric)
p	The effect size, i.e., the probability that the first random variable is less than the second random variable (P(X<Y)) (numeric)
alpha	Type I error rate or significance level (numeric)
distn	Base R's name for the first distribution (known as X in the above notation) and any required parameters. Supported distributions are normal, exponential, and double exponential ("norm","exp", "doublex"). User may enter distribution without parameters, and default parameters will be set (i.e., "norm" defaults to "norm(0,1)"), or user may specify both distribution and parameters (i.e., "norm(0,1)").
sides	Options are "two.sided", "less", or "greater". "less" means the alternative hypothesis is that distn is less than distm (string)
k	Standard deviation (SD) scalar for use with the normal or double exponential distribution options. The SD for distm is computed as k multiplied by the SD for distn. Equivalently, k is the ratio of the SDs of the second and first distribution (k = SDm/SDn). Default is k=1 (equal SDs) (numeric)
wmwodds	The effect size expressed as odds = p/(1-p). Either p or wmwodds must be input (numeric)
nsims	Number of simulated datasets for calculating power; 10,000 is the default. For exact power to the hundredths place (e.g., 0.90 or 90%) around 100,000 simulated datasets is recommended (numeric)

References

Mollan K.R., Trumble I.M., Reifeis S.A., Ferrer O., Bay C.P., Baldoni P.L., Hudgens M.G. Exact Power of the Rank-Sum Test for a Continuous Variable, arXiv:1901.04597 [stat.ME], Jan. 2019.

Examples

```
# We want to calculate the statistical power to compare the distance between mutations on a DNA
# strand in two groups of people. Each group (X and Y) has 10 individuals. We assume that the
# distance between mutations in the first group is exponentially distributed with rate 3. We assume
# that the probability that the distance in the first group is less than the distance in the second
# group (i.e., P(X<Y)) is 0.8. The desired type I error is 0.05.

wmwpowp(n = 10, m = 10, distn = "exp(3)", p = 0.8, sides = "two.sided", alpha = 0.05)
```

Index