Package ‘wmwAUC’

May 8, 2026

Title Wilcoxon-Mann-Whitney Test of No Group Discrimination

Version 0.2.0

Date 2025-12-07

Description Provides inference for the Wilcoxon-Mann-Whitney test under the null hypothesis H0: AUC = 0.5 for continuous, discrete or mixed random variables. Traditional implementations test H0: F = G, which is inappropriately broad and leads to erroneous inferences. Methods are described in M. Grendar (2025) ``Wilcoxon-Mann-Whitney Test of No Group Discrimination'' <doi:10.48550/arXiv.2511.20308>.

License MIT + file LICENSE

URL `https://github.com/grendar/wmwAUC`

BugReports `https://github.com/grendar/wmwAUC/issues`

Encoding UTF-8

RoxygenNote 7.3.2

Depends R (>= 4.0.0)

Suggests testthat (>= 3.0.0), ggsci, viridis, gemR, gss, knitr, rmarkdown, ggbeeswarm, ggplot2, qqplotr, rlang, twosamples, patchwork, sfsmisc, stats, VGAM

Config/testthat/edition 3

NeedsCompilation no

Author Marian Grendar [aut, cre] (ORCID: <`https://orcid.org/0000-0002-6712-3457`>)

Maintainer Marian Grendar <`marian.grendar@gmail.com`>

Repository CRAN

Date/Publication 2025-12-19 14:20:02 UTC

Contents

`Ex2` *Synthetic data*

Description

A data frame with numeric `y` and factor `group`

Usage

```
data(Ex2)
```

Format

A data frame with 200 observations on 2 variables.

`plot.wmw_test` *Plot Method for wmw_test Objects*

Description

Creates empirical ROC curve plot with test results (p-value, eAUC with confidence interval) displayed in subtitle. If `ci_method = 'boot'` was used in `wmw_test()`, the plot includes confidence bands for the ROC curve constructed using the same bootstrap resamples used for the AUC confidence interval.

Usage

```
## S3 method for class 'wmw_test'
plot(x, combine_plots = TRUE, ...)
```

Arguments

`x`	Object of class 'wmw_test' returned by `wmw_test()`
`combine_plots`	Logical, whether to return combined plot using patchwork (TRUE) or list of individual plots (FALSE). Only relevant when special_case = TRUE
`...`	Additional arguments (not currently used)

Details

When `special_case = TRUE` was used in `wmw_test()`, an additional boxplot with swarmplot overlay is created, showing the eAUC as effect size estimate with confidence interval in the subtitle (demonstrating the dual interpretation of eAUC in the location-shift case).

Value

No return value, called for side effects. Creates a plot visualizing the Wilcoxon-Mann-Whitney test results including distributions, test statistic, and confidence information.

`print.wmw_test` *Print Method for wmw_test Objects*

Description

Prints summary of Wilcoxon-Mann-Whitney discrimination test results.

Usage

```
## S3 method for class 'wmw_test'
print(x, digits = 3, ...)
```

Arguments

`x`	Object of class 'wmw_test' returned by `wmw_test()`
`digits`	Integer, number of digits to display for numeric results (default: 4)
`...`	Additional arguments (not currently used)

Value

Invisibly returns the input object `x` (of class "wmw_test"). Called primarily for side effects to print a formatted summary of the Wilcoxon-Mann-Whitney test results to the console.

`pseudomedian_ci` *Confidence Interval for Hodges-Lehmann Pseudomedian via Test Inversion*

Description

Computes confidence interval for the pseudomedian under $\mathrm{H}_0: \mathrm{AUC} = 0.5$ by test inversion.

Usage

```
pseudomedian_ci(x, y, conf.level = 0.95, pvalue_method = "EU", n_grid = 1000)
```

Arguments

`x`	numeric vector, first sample
`y`	numeric vector, second sample
`conf.level`	confidence level (default 0.95)
`pvalue_method`	character, either 'EU' or 'BC'
`n_grid`	number of grid points for search (default 1000)

Value

list with conf.int, estimate and conf.level

`quadruplot` *Four EDA Plots for Visual Assessment of Location-Shift Assumption*

Description

Creates four diagnostic plots to visually assess whether the location-shift assumption $F_1(x) = F_2(x - \delta)$ holds: (1) boxplot with swarmplot overlay, (2) density plot comparison, (3) wormplot of median-centered residuals, and (4) empirical CDF comparison with confidence band for median-centered data.

Usage

```
quadruplot(
  formula,
  data,
  ref_level = NULL,
  test = "ks",
  seed = 123L,
  ylab = NULL,
  color_palette = "lancet",
  combine_plots = TRUE,
  distribution = "norm",
  show_colors = TRUE,
  show_legend = TRUE
)
```

Arguments

`formula`	Formula of the form `response ~ group`
`data`	Data frame containg response, group
`ref_level`	Character, reference level of the grouping factor. If NULL (default), uses first factor level

`test`	Character, statistical test for shift-equivalence assumption. Tests for distributional equality applied to median-centered data: "ks" (Kolmogorov-Smirnov) (default), "kuiper" (Kuiper), "cvm" (Cramér-von Mises), "ad" (Anderson-Darling), "wass" (Wasserstein), "dts" (DTS test).
`seed`	Numeric, for set.seed() used in `test_shift_equivalence()` for bootstrap.
`ylab`	Character, label for y-axis. If NULL (default), uses variable name
`color_palette`	Character, color palette to use. One of "viridis", "plasma", "inferno", "magma", or "cividis"
`combine_plots`	Logical, whether to return combined plot using patchwork (TRUE) or list of individual plots (FALSE)
`distribution`	Character, theoretical distribution for Q-Q plot comparison. Default is "norm" for normal distribution
`show_colors`	Logical, whether to use colors (TRUE) or grayscale (FALSE)
`show_legend`	Logical, whether to display legend in plots (default TRUE)

Details

The location-shift assumption is assessed by applying a test of H0: equality of distributions to median-centered data. One of the tests from the `twosamples` package can be used. The empirical CDF plot includes 95% confidence bands for the difference between distributions, computed using the `sfsmisc::KSd` function based on the Kolmogorov-Smirnov distribution. These bands help assess whether observed differences between median-centered distributions exceed what would be expected under the location-shift assumption.

Value

If combine_plots = TRUE, returns a combined ggplot object created by patchwork. If FALSE, returns a list of four ggplot objects named 'boxplot', 'density', 'wormplot', and 'ecdf'.

Note

Uses **twosamples** for distribution comparison and `KSd` from **sfsmisc** for exact confidence bands.

References

O'Dowd, C. (2025). Statistical Code Examples. https://codowd.com/code (accessed November 28, 2025).

Maechler M (2024). *sfsmisc: Utilities from 'Seminar fuer Statistik' ETH Zurich*. R package version 1.1-20, https://CRAN.R-project.org/package=sfsmisc.

Examples

```
library(wmwAUC)

data(Ex2)
da <- Ex2
qp = quadruplot(y ~ group, data = da, ref_level = 'control')
qp
```

`simulation1` *Synthetic data*

Description

Synthetic data

Usage

```
data(simulation1)
```

Format

A list containing simulation results (N=10000, n=1000):

eauc Empirical AUC values

pval_wt Traditional wilcox.test p-values

pval_wmw WMW p-values under H0: AUC = 0.5

`simulation2` *Synthetic data*

Description

Synthetic data

Usage

```
data(simulation2)
```

Format

A list containing simulation results (N=10000, n=1000):

eauc Empirical AUC values

pval_wt Traditional wilcox.test p-values

pval_wmw WMW p-values under H0: AUC = 0.5

simulation3 *Synthetic data*

Description

Synthetic data

Usage

```
data(simulation3)
```

Format

A list containing simulation results (N=500, n=300):

wmw_ci 95% confidence intervals obtained by pseudomedian_ci()

wt_ci 95% confidence intervals obtained by wilcox.test()

eauc Values of eAUC

pseudomedian Values of the pseudomedian

wmw_pvalue *P-value for Wilcoxon-Mann-Whitney Test of No Group Discrimination (Continuous Variables)*

Description

Tests $\mathrm{H}_0: \mathrm{AUC} = 0.5$ vs $\mathrm{H}_1: \mathrm{AUC} \neq 0.5$ with proper finite-sample corrections

Usage

```
wmw_pvalue(x, y, alternative = "two.sided")
```

Arguments

x	Numeric vector of cases/group 1 values
y	Numeric vector of controls/reference group values
alternative	character: "two.sided", "greater", or "less"

Details

Implements the Bias-Corrected (BC) variance estimator with second-order U-statistic correction to provide honest p-values under $\mathrm{H}_0: \mathrm{AUC} = 0.5$. Uses three-tier approach: permutation $(n < 20)$, bias-corrected $(20 \leq n < 50)$, asymptotic with correction $n \geq 50$.

For medium samples, the naive variance estimators $\widehat{\mathrm{Var}}(G(X))$ and $\widehat{\mathrm{Var}}(F(Y))$ are corrected by subtracting O(1/n) bias terms of the form $(n_1 n_2)^{-1} \sum_i \hat{G}(X_i)(1 - \hat{G}(X_i))$ to prevent variance underestimation that would inflate Type I error rates.

Function assumes x represents cases and y represents the reference level, in accord with `wilcox.test()` and `wmw_test()`. Internal calculations convert to P(X < Y) framework to match theoretical derivations.

Value

p-value

`wmw_pvalue_ties` *P-value for Wilcoxon-Mann-Whitney Test of No Group Discrimination (With Possible Ties)*

Description

Tests $\mathrm{H}_0: \mathrm{AUC} = 0.5$ vs $\mathrm{H}_1: \mathrm{AUC} \neq 0.5$ with exact finite-sample unbiased variance estimation for arbitrary tie patterns

Usage

```
wmw_pvalue_ties(x, y, alternative = "two.sided")
```

Arguments

`x`	Numeric vector of cases/group 1 values
`y`	Numeric vector of controls/reference group values
`alternative`	character: "two.sided", "greater", or "less"

Details

Implements the Exact finite-sample Unbiased (EU) variance estimator derived from Hoeffding decomposition theory. Uses tie-corrected kernel $h(x, y) = \mathbf{1}\{x < y\} + \frac{1}{2}\mathbf{1}\{x = y\}$ with universal second-order correction factor to provide honest p-values under $\mathrm{H}_0: \mathrm{AUC} = 0.5$ regardless of tie structure.

Uses three-tier approach: permutation $(n < 20)$, exact unbiased estimator $(20 \leq n < 50)$, asymptotic with corrections $n \geq 50$.

The unbiased variance estimator is constructed as a specific linear combination:

$\widetilde{\mathrm{Var}}(\hat{A}) = \frac{n_2 \hat{\zeta}_1^2 + n_1 \hat{\zeta}_2^2 - \frac{M-1}{M}\hat{v}}{M+1}$

where $\hat{v}$ is the pooled sample variance of kernel values and $\hat{\zeta}_1^2, \hat{\zeta}_2^2$ are row/column mean variances.

Welch-Satterthwaite degrees of freedom account for bias correction structure:

$$\nu = \frac{(\hat{\sigma}^2)^2}{\frac{(n_2\hat{\zeta}_1^2/(M+1))^2}{n_1-2} + \frac{(n_1\hat{\zeta}_2^2/(M+1))^2}{n_2-2} + \frac{((M-1)\hat{v}/(M(M+1)))^2}{M-3}}$$

Function uses mid-rank tie handling throughout, ensuring theoretical consistency with the corrected null hypothesis framework.

Function assumes x represents cases and y represents the reference level, in accord with `wilcox.test()` and `wmw_test()`. Internal calculations convert to P(X < Y) framework to match theoretical derivations.

Value

p-value

`wmw_test` *Wilcoxon-Mann-Whitney Test of No Group Discrimination*

Description

Performs distribution-free Wilcoxon-Mann-Whitney test for AUC-detectable group discrimination, testing $\mathrm{H}_0: \mathrm{AUC} = 0.5$ against $\mathrm{H}_1: \mathrm{AUC} \neq 0.5$. Under location-shift assumption, equivalently tests zero location difference.

Usage

```
wmw_test(
  formula,
  data,
  ref_level = NULL,
  special_case = FALSE,
  alternative = c("two.sided", "greater", "less"),
  pvalue_method = "EU",
  ci_method = "hanley",
  conf_level = 0.95,
  n_grid = 100,
  ...
)
```

Arguments

`formula`	Formula of the form `response ~ group`
`data`	Data frame containing continuous response variable and grouping factor
`ref_level`	Character, reference level of grouping factor (if NULL, uses first level)
`special_case`	Logical, location-shift assumption (default FALSE)
`alternative`	Character, alternative hypothesis is c("two.sided", "greater", "less")

pvalue_method	Character, method ('EU', 'BC') used for computing p-values; 'BC' assumes continuous data (default 'EU')
ci_method	Character, confidence interval method for eAUC: c('hanley', 'boot', 'none')
conf_level	Numeric, confidence level for intervals (default 0.95)
n_grid	Numeric, number of grid points for search in `pseudomedian_ci()` (default 100)
...	Additional arguments passed to `roc_with_ci()`

Details

The function tests the null hypothesis $\mathrm{H_0: AUC} = 0.5$ against $\mathrm{H_0: AUC} \neq 0.5$, where AUC represents the Area Under the ROC Curve and - following the convention of `wilcox.test()` - equals the probability $P(X > Y)$ that a randomly selected observation from the first group exceeds a randomly selected observation from the second group.

For `response ~ group`, observations from the non-reference group constitute X, while observations from the reference group (specified by `ref_level`) constitute Y. Thus AUC = P(non-reference > reference). If `ref_level` is not specified, the first factor level is used as reference. The U statistic and the resulting empirical AUC (eAUC) are calculated consistently with this group assignment.

The test statistic is eAUC, which estimates the true AUC. The empirical ROC curve (eROC) is constructed by varying the classification threshold across all observed values and computing sensitivity and 1-specificity at each threshold.

When special_case = TRUE, the function additionally reports location-shift parameters under the assumption that $F_1(x) = F_2(x-\delta)$. Under this assumption, the discrimination test $\mathrm{H_0: AUC} = 0.5$ is mathematically equivalent to testing H0: $\delta = 0$ (zero location shift). In this special case, eAUC takes the dual role of both test statistic and effect size for the location difference.

Confidence intervals for the true AUC are computed using either the Hanley and McNeil (1982) method based on asymptotic normality, or bootstrap resampling. If bootstrap resampling is selected, it is also used for constructing the confidence band for the ROC curve.

The function uses `Exact Unbiased` ('EU') method for computing p-values that can handle any type of data (continuous, discrete, mixed). The `Bias-Corrected` ('BC') method that requires continuous data is provided through `pvalue_method = 'BC'` option.

Constructs confidence intervals for the pseudomedian via test inversion. Under location-shift assumptions ($G(x) = F(x-\delta)$), the pseudomedian represents the location difference between groups.

Statistical Methodology: Unlike standard implementations that assume the erroneously broad null hypothesis $\mathrm{H_0: F = G}$, this function derives p-values under the correct null hypothesis $\mathrm{H_0: AUC} = 0.5$ that WMW actually tests. P-values are computed using asymptotic distribution theory with two methods of finite-sample bias corrections:

1. Exact Unbiased ('EU') estimation of variance of eAUC which handles any type of data (continuous, discrete, mixed);
2. Bias Correction ('BC') sample-size dependent method to maintain proper Type I error control. Confidence intervals for the pseudomedian are obtained by inverting the test.

Value

Object of class 'wmw_test' containing:

`special_case` Logical indicating whether special case (location-shift) analysis was performed

`n` Named vector with components n1, n2 giving sample sizes for each group

`U_statistic` U statistic

`p_value` P-value for testing H0: AUC = 0.5

`alternative` Alternative hypothesis specification

`pvalue_method` Character string describing the test method

`data_name` Character string giving the name of the data

`pseudomedian` Hodges-Lehmann median difference estimate (when special_case = TRUE)

`pseudomedian_conf_int` Confidence interval for the location shift (when special_case = TRUE)

`pseudomedian_conf_level` Confidence level for the confidence interval for HL estimator (when special_case = TRUE)

`ci_method` Method used to compute confidence interval for AUC

`roc_object` ROC analysis object returned by `roc_with_ci` function

`auc` Empirical AUC (eAUC), the standardized U statistic

`auc_conf_int` Confidence interval for true AUC using Hanley-McNeil or bootstrap method

`x_vals` Numeric vector of observations from non-reference group

`y_vals` Numeric vector of observations from reference group

`groups` Character vector of group labels from original data

`group_levels` Character vector of factor levels for grouping variable

`group_ref_level` Character string indicating which level corresponds to reference group

References

Wilcoxon, F. (1945). Individual comparisons by ranking methods. Biometrics Bulletin, 1(6), 80-83.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. The Annals of Mathematical Statistics, 18(1), 50-60.

Van Dantzig, D. (1951). On the consistency and the power of Wilcoxon's two sample test. Proceedings KNAW, Series A, 54(1), 1-8.

Birnbaum, Z. W. (1956). On a use of the Mann-Whitney statistic. In Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics (Vol. 3, pp. 13-18). University of California Press.

Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. Journal of mathematical psychology, 12(4), 387-415.

Lehmann, E. L., & Abrera, H. B. D. (1975). Nonparametrics. Statistical methods based on ranks. San Francisco, CA, Holden-Day.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 143(1), 29-36.

Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. Psychological bulletin, 114(3), 494.

Arcones, M. A., Kvam, P. H., & Samaniego, F. J. (2002). Nonparametric estimation of a distribution subject to a stochastic precedence constraint. Journal of the American Statistical Association, 97(457), 170-182.

Pepe, M. S. (2003). The statistical evaluation of medical tests for classification and prediction. Oxford university press.

Conroy, R. M. (2012). What hypotheses do “nonparametric” two-group tests actually test?. The Stata Journal, 12(2), 182-190.

del Barrio, E., Cuesta-Albertos, J. A., & Matrán, C. (2025). Invariant measures of disagreement with stochastic dominance. The American Statistician, 1-13.

Grendar, M. (2025). Wilcoxon-Mann-Whitney test of no group discrimination. arXiv:2511.20308.

See Also

print.wmw_test for formated output of wmw_test(). plot.wmw_test for plot of output of wmw_test(). wmw_pvalue for details on computing p-values in the continuous case ('BC') wmw_pvalue_ties for details on computing p-values in the 'EU' mode pseudomedian_ci for details on computing confidence intervals for pseudomedian quadruplot for exploratory data analysis plots that assist in evaluating location-shift assumption validity. wilcox.test for Wilcoxon-Mann-Whitney test in base R.

Examples

```
library('wmwAUC')
# Ex 1

library('gemR')
data(MS)
da <- MS
# preparing data frame
class(da$proteins) <- setdiff(class(da$proteins), "AsIs")
df <- as.data.frame(da$proteins)
df$MS <- da$MS
# WMW test
wmd <- wmw_test(P19099 ~ MS, data = df, ref_level = 'no')
wmd
plot(wmd)
# EDA to assess location shift assumption validity
qp <- quadruplot(P19099 ~ MS, data = df, ref_level = 'no')
qp
# => location shift assumption is not valid


# Ex 2

data(Ex2)
da <- Ex2
# WMW test
wmd <- wmw_test(y ~ group, data = da, ref_level = 'control')
wmd
plot(wmd)
```

```
# Check location-shift assumption with EDA
qp <- quadruplot(y ~ group, data = da, ref_level = 'control', test = 'ks')
qp
# => location-shift assumption not tenable
# Note that medians are essentially the same:
median(da$y[da$group == 'case'])
# 0.495
median(da$y[da$group == 'control'])
# 0.493
# Erroneous use of location-shift special case would falsely
# conclude significant median difference despite identical medians
wml <- wmw_test(y ~ group, data = da, special_case = TRUE,
                ref_level = 'control')
wml


# Ex 3

library('gss')
data(wesdr)
da = wesdr
da$ret = as.factor(da$ret)
# WMW
wmd <- wmw_test(bmi ~ ret, data = da, ref_level = '0')
wmd
plot(wmd)
# EDA to assess location shift assumption validity
qp <- quadruplot(bmi ~ ret, data = da, ref_level = '0')
qp
# => location shift assumption is tenable
# Special case of WMW test
wml <- wmw_test(bmi ~ ret, data = da, ref_level = '0',
                ci_method = 'boot', special_case = TRUE)
wml
plot(wml)
```

Index