

# Package: testforDEP (via r-universe)

October 10, 2024

**Type** Package

**Title** Dependence Tests for Two Variables

**Version** 0.2.0

**Author** Jeffrey C. Miecznikowski, En-shuo Hsu, Yanhua Chen, Albert Vexler

**Maintainer** En-shuo Hsu <daviden1013@gmail.com>

**Description** Provides test statistics, p-value, and confidence intervals based on 9 hypothesis tests for dependence.

**License** GPL-3

**LazyData** TRUE

**Imports** Rcpp (>= 0.12.7), methods

**Depends** R (>= 3.2.5), parallel, minerva, Hmisc

**LinkingTo** Rcpp

**RoxygenNote** 5.0.1

**NeedsCompilation** yes

**Date/Publication** 2017-01-20 10:49:22

**Repository** <https://daviden1013.r-universe.dev>

**RemoteUrl** <https://github.com/cran/testforDEP>

**RemoteRef** HEAD

**RemoteSha** c646f49da6b8495df5e04e3d5d7e7282c84e666e

## Contents

AUK . . . . .	2
EL . . . . .	3
Hoeffding . . . . .	3
Kallenberg . . . . .	4
Kendall . . . . .	4
LSAT . . . . .	4
MIC . . . . .	5

Pearson . . . . .	5
Spearman . . . . .	5
testforDEP . . . . .	6
Vexler . . . . .	7

<b>Index</b>	<b>8</b>
--------------	----------

---

AUK	<i>Draw Kendall plot and compute AUK.</i>
-----	---

---

### Description

This function draws Kendall plot of 2 variables. Also provides an index AUK (area under Kendall plot).

### Usage

```
AUK(x, y, plot = F, main = "Kendall plot", Auxiliary.line = T,
    BS.CI = 0, set.seed = FALSE)
```

### Arguments

<code>x</code>	a numeric vector stores first variable.
<code>y</code>	a numeric vector stores second variable.
<code>plot</code>	a TRUE/ FALSE flag for generating Kendall plot or not.
<code>main</code>	a character indicating the title of the plot.
<code>Auxiliary.line</code>	a TRUE/ FALSE flag for drawing auxiliary lines or not.
<code>BS.CI</code>	a numeric specifying alpha for Bootstrap confidence interval. When equal 0, confidence interval won't be computed.
<code>set.seed</code>	a TRUE/ FALSE flag specifying setting seed or not.

### Details

AUK is bounded between 0 and 0.75. For positively correlated  $x$  and  $y$ 's, say  $x = y$ ,  $AUK = 0.75$ . And the plot follows the concave auxiliary line. While negatively correlated  $x$  and  $y$ 's,  $AUK = 0$ . The plot is horizontal on  $y = 0$ . For independent  $x$  and  $y$ ,  $AUK = 0.5$ . Kendall plot is on the diagonal. Due to possible variable overflow, this function is only suitable for input size less than 1000. Input size greater than 1000 causes error.

### Value

a list containing a numeric AUK, a numeric vector `W.in` ( $x$  axis of plot), a numeric vector `Hi.sort` ( $y$  axis of plot), and three confidence intervals: normal CI, pivotal CI and percentage CI.

### Author(s)

Jeffrey C. Miecznikowski, En-shuo Hsu, Yanhua Chen, Albert Vexler

**References**

Vexler, Albert, Xiwei Chen, and Alan D. Hutson. "Dependence and independence: Structure and inference." *Statistical methods in medical research* (2015): 0962280215594198.

R package "VineCopula": Schepsmeier, Ulf, et al. "Package 'VineCopula'." (2015).

**Examples**

```
set.seed(123)
x = runif(100)
y = runif(100)

result = AUK(x, y, plot = TRUE)
result$AUK

#[1] 0.4987523
```

---

EL

*Empirical Likelihood based test for dependence*

---

**Description**

Empirical Likelihood based test for dependence. See references.

**References**

Einmahl, J. H., & McKeague, I. W. (2003). Empirical likelihood based hypothesis testing. *Bernoulli*, 267-290.

---

Hoeffding

*Hoeffding's test for dependence*

---

**Description**

Test statistic is computed by `hoeffd{Hmisc}`. See [hoeffd](#). Note that test statistic D is 30 times the original test statistic in the original publication.

**References**

Harrell Jr FE, Dupont MC (2006). "The Hmisc Package." R package version, 3, 0-12.

---

Kallenberg	<i>Kallenberg test for dependence</i>
------------	---------------------------------------

---

**Description**

Includes TS2 and V. See reference.

**References**

Kallenberg WC, Ledwina T (1999). Data-Driven Rank Tests for Independence." 94. doi: 10.1080/01621459.1999.10473844.

---

Kendall	<i>Kendall test for dependence</i>
---------	------------------------------------

---

**Description**

Test statistic is computed by `cor.test{stats}`. See [cor.test](#). Note that test statistic returned is the pivot  $z$  that approximately follows normal distribution.

---

LSAT	<i>LSAT dataset</i>
------	---------------------

---

**Description**

A dataset of average law school admission test (LSAT) and grade point average (GPA) from 82 American law schools participated in a large study of admission practices.

**Usage**

```
data("LSAT")
```

**Format**

A data frame with 82 observations on the following 3 variables.

School a numeric vector of school numbers.

LSAT a numeric vector of LSAT's.

GPA a numeric vector of GPA's.

**Details**

details see references.

**Source**

Efron B, Tibshirani RJ (1994). An Introduction to the Bootstrap. CRC Press.

**References**

Efron B, Tibshirani RJ (1994). An Introduction to the Bootstrap. CRC Press.

---

MIC	<i>MIC test for dependence</i>
-----	--------------------------------

---

**Description**

Test statistic is computed by `mine{minerva}`. See [mine](#).

---

Pearson	<i>Pearson test for dependence</i>
---------	------------------------------------

---

**Description**

Pearson test for linear dependence. Note that test statistic returned is the pivot  $t$  that follows Student's  $t$  distribution.

---

Spearman	<i>Spearman test for dependence</i>
----------	-------------------------------------

---

**Description**

Test statistic is computed by `cor.test{stats}`. See [cor.test](#). Note that test statistic returned is the pivot  $t$  that approximately follows Student's  $t$  distribution. Spearman test cannot handle tie. Since bootstrap resamples with replacement which generates ties, bootstrap confidence interval does not apply. Setting `BS.CI > 0` throughs warning message.

testforDEP

*Test dependence for two data***Description**

This function computes test statistic, p value, and confidence interval for dependence based on classic methods: Pearson, Kendall, Spearman, and modern methods: Vexler, Kallenberg, MIC, Hoeffding, and Empirical Likelihood tests.

**Usage**

```
testforDEP(x = NA, y = NA, data = NA, test, p.opt = "MC",
  num.MC = 10000, BS.CI = 0, rm.na = FALSE, set.seed = FALSE)
```

**Arguments**

x	a numeric vector stores first variable.
y	numeric vector stores second variable.
data	(Optional) a data frame stores data to be tested.
test	a character indicating which test to implement.. Must be one of {"PEARSON", "KENDALL", "SPEARMAN", "VEXLER", "TS2", "V", "MIC", "HOEFFD", "EL"}
p.opt	a character specifying p value to be obtained by distribution or by Monte Carlo simulation. Must be "dist", "MC" or "table".
num.MC	a numeric for number of Monte Carlo simulations.
BS.CI	a numeric specifying alpha for Bootstrap confidence interval. When equal 0, confidence interval won't be computed.
rm.na	a TRUE/ FALSE flag indicating whether remove missing data (NA) in input.
set.seed	a TRUE/ FALSE flag indicating whether set seed for Monte Carlo simulation and bootstrap sampling.

**Details**

Argument "x, y" and "data" are two different ways to input data. When x or y is missing, data will be taken as input; while x, y and data all exist leads to error. Argument data is a two-column numeric data frame. The order of columns does not affect results. Since modern test methods: "VEXLER", "TS2", "V", "MIC", "HOEFFD", and "EL" have no continuous probability density function, argument p.opt = "dist" does not apply. For classic methods, when p.opt is "dist", argument num.MC will be ignored. p.opt = "table" use interpolation from pre stored simulated tables. Current version only supports "VEXLER", "MIC", "HOEFFD" and "EL" tests. For Vexler, MIC and EL, since computation is more time-consuming, a warning with estimated execution time will be returned when input size > 100. Input size <= 100 is recommended for Monte Carlo p-value. For input size > 100 use table. num.MC should be a integer between 100 and 10,000 for acceptable computation times. NA in input is not acceptable. Set rm.na = TRUE to remove. More details see [Pearson](#), [Kendall](#), [Spearman](#), [Vexler](#), [Kallenberg](#), [MIC](#), [Hoeffding](#), [EL](#).

**Value**

an S4 object of class "testforDEP\_result", having attributes: test statistics (TS), p value (p\_value) and confidence interval (CI) if apply.

**Author(s)**

Jeffrey C. Miecznikowski, En-shuo Hsu, Yanhua Chen, Albert Vexler

**See Also**

Technical report: <http://sphhp.buffalo.edu/content/dam/sphhp/biostatistics/Documents/techreports/UB-Biostatistics-TR1701.pdf>

**Examples**

```
set.seed(123)
x = runif(100, 0, 1)
y = runif(100, 0, 1)

testforDEP(x, y, test = "SPEARMAN", p.opt = "MC",
           num.MC = 10000, BS.CI = 0, set.seed = TRUE)

#An object of class "testforDEP_result"
#Slot "TS":
#[1] 59.54311

#Slot "p_value":
#[1] 0.6735326

#Slot "CI":
#list()
```

---

Vexler

*Vexler's test for dependence*

---

**Description**

A method based on empirical likelihood ratio test. Published by Dr. Vexler in 2014. See reference.

**References**

Vexler A, Tsai WM, Hutson AD (2014). A Simple Density-Based Empirical Likelihood Ratio Test for Independence."

# Index

\* **datasets**

LSAT, 4

AUK, 2

cor.test, 4, 5

EL, 3, 6

hoeffd, 3

Hoeffding, 3, 6

Kallenberg, 4, 6

Kendall, 4, 6

LSAT, 4

MIC, 5, 6

mine, 5

Pearson, 5, 6

Spearman, 5, 6

testforDEP, 6

Vexler, 6, 7