Package 'ircor'

October 13, 2022

Title Correlation Coefficients for Information Retrieval

Version 1.0

Description Provides implementation of various correlation coefficients of common use in
Information Retrieval. In particular, it includes Kendall (1970, isbn:0852641990) tau coefficient
as well as tau_a and tau_b for the treatment of ties. It also includes Yilmaz et al. (2008)
<doi:10.1145/1390334.1390435> tauAP correlation coefficient, and ver-
sions tauAP_a and tauAP_b
developed by Urbano and Marrero (2017) <doi:10.1145/3121050.3121106> to cope with ties.

Depends R (>= 3.2.0)

License MIT + file LICENSE

BugReports https://github.com/julian-urbano/ircor/issues

URL https://github.com/julian-urbano/ircor/

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

Suggests testthat

NeedsCompilation no

Author Julián Urbano [aut, cre],
Mónica Marrero [aut]

Maintainer Julián Urbano <urbano.julian@gmail.com>

Repository CRAN

Date/Publication 2017-08-21 09:07:29 UTC

R topics documented:

tau *Kendall τ Rank Correlation Coefficients*

Description

`tau` is the rank correlation coefficient by Kendall, where neither vector can contain tied items. `tau_a` and `tau_b` are the versions developed to cope with ties under the scenarios of accuracy and agreement, respectively. See the references for details.

Usage

```
tau(x, y)

tau_a(x, y)

tau_b(x, y)
```

Arguments

x   a numeric vector. In `tau_a` this is the vector of true scores.

y   a numeric vector of the same length as x. In `tau_a` this is the vector of estimated scores.

Value

The correlation coefficient.

References

M.G. Kendall (1970). Rank Correlation Methods. Charles Griffin & Company Limited.

See Also

tauAP for AP correlation coefficients.

Examples

```
# No ties
x <- c(0.67, 0.45, 0.29, 0.12, 0.57, 0.24, 0.94, 0.75, 0.08, 0.54)
y <- c(0.48, 0.68, 0.32, 0.09, 0.06, 0.61, 0.87, 0.22, 0.44, 0.84)
tau(x, y)
tau_a(x,y) # same as tau
tau_b(x,y) # same as tau

# Ties in y
y <- round(y, 1)
tau_a(x, y)
tau_b(x, y)
```

```
# Ties in x too
x <- round(x, 1)
tau_b(x, y)
```

---

tauAP   *AP Rank Correlation Coefficients*

Description

`tauAP` is the AP rank correlation coefficient by Yilmaz et al., where neither vector can contain tied items. `tauAP_a` and `tauAP_b` are the versions developed by Urbano and Marrero to cope with ties under the scenarios of accuracy and agreement, respectively. See the references for details.

Usage

```
tauAP(x, y, decreasing = TRUE)

tauAP_a(x, y, decreasing = TRUE)

tauAP_b(x, y, decreasing = TRUE)
```

Arguments

x	a numeric vector. In `tauAP_a` this is the vector of true scores.
y	a numeric vector of the same length as x. In `tauAP_a` this is the vector of estimated scores.
decreasing	logical. Should the sort order be increasing or decreasing (default)?

Details

Note that the sorting order is decreasing by default, as should be for instance if the scores represent the effectiveness of systems. When the sorting order is ascending, as is for instance when the vectors represent ranks, the parameter `decreasing` must be set to `FALSE`.

Value

The correlation coefficient.

References

E. Yilmaz, J.A. Aslam and S. Robertson (2008). A New Rank Correlation Coefficient for Information Retrieval. ACM SIGIR.

J. Urbano and M. Marrero (2017). The Treatment of Ties in AP Correlation. ACM ICTIR.

See Also

[tau](#) for Kendall correlation coefficients.

Examples

```
# No ties
x <- c(0.67, 0.45, 0.29, 0.12, 0.57, 0.24, 0.94, 0.75, 0.08, 0.54)
y <- c(0.48, 0.68, 0.32, 0.09, 0.06, 0.61, 0.87, 0.22, 0.44, 0.84)
tauAP(x, y)
tauAP_a(x,y) # same as tauAP

# Ties in y
y <- round(y, 1)
tauAP_a(x, y)
tauAP_b(x, y)

# Ties in x too
x <- round(x, 1)
tauAP_b(x, y)

# Set decreasing to FALSE when x and y already represent ranks
x <- rank(-x)
y <- rank(-y)
tauAP_b(x, y, FALSE) # same as above
```

Index