Package ‘gt4ireval’

October 13, 2022

Title Generalizability Theory for Information Retrieval Evaluation
Version 2.0
Description Provides tools to measure the reliability of an Information Retrieval test collection. It allows users to estimate reliability using Generalizability Theory and map those estimates onto well-known indicators such as Kendall tau correlation or sensitivity.
Depends R (>= 3.2)
License MIT + file LICENSE
BugReports https://github.com/julian-urbano/gt4ireval/issues
URL https://github.com/julian-urbano/gt4ireval/
Encoding UTF-8
LazyData true
Suggests testthat, knitr, rmarkdown
RoxygenNote 6.0.1
VignetteBuilder knitr
NeedsCompilation no
Author Julián Urbano [aut, cre]
Maintainer Julián Urbano <urbano.julian@gmail.com>
Repository CRAN
Date/Publication 2017-03-06 08:29:02

R topics documented:

adhoc3 ................................................. 2
dstudy ................................................... 2
gstudy .................................................... 4
gt2tau .................................................... 5
synthetic4 .............................................. 6

Index 7
Description

This is the set of Average Precision scores of the 40 systems submitted to the TREC-3 Ad hoc track, evaluated over 50 topics.

Usage

adhoc3

Format

A data frame with 40 columns (systems) and 50 rows (queries).

References


See Also

http://trec.nist.gov

dstudy

D-study (Decision)

Description

dstudy runs a D-study from the results of a gstudy and computes, for a certain number of queries, the expected generalizability coefficient Erho2 and index of dependability Phi, possibly with confidence intervals. Alternatively, it can estimate the number of queries needed to achieve a certain level of stability, also with confidence intervals.

Usage

dstudy(gdata, queries = gdata$n.q, stability = 0.95, alpha = 0.025)

Arguments

gdata The result of running a gstudy with existing data.
queries A vector with different query set sizes for which to estimate Erho2 and Phi. Defaults to the number of queries used to compute gdata.
stability A vector with target Erho2 and Phi values to estimate required query set sizes.
alpha A vector of confidence levels to compute intervals for Erho2, Phi and query set sizes. This is the probability on each side of the interval, so for a 90% confidence interval one must set alpha to 0.05.
Value

An object of class `dstudy`, with the following components:

- `Erho2`, `Erho2.lwr`, `Erho2.upr` Expected generalizability coefficient, and lower and upper limits of the intervals around it.
- `Phi`, `Phi.lwr`, `Phi.upr` Expected index of dependability, and lower and upper limits of the intervals around it.
- `n.q_Erho2`, `n.q_Erho2.lwr`, `n.q_Erho2.upr` Expected number of queries to achieve the generalizability coefficient, and lower and upper limits of the intervals around it.
- `n.q_Phi`, `n.q_Phi.lwr`, `n.q_Phi.upr` Expected number of queries to achieve the index of dependability, and lower and upper limits of the intervals around it.
- `call` A list with the `gstudy` used in this D-study, the target number of queries, target level of stability and `alpha` level for the confidence intervals.

Author(s)

Julián Urbano

References


See Also

`gstudy, gt2tau`

Examples

```r
# estimate stability at various query set sizes
g <- gstudy(adhoc3)
dstudy(g)
# estimate required query set sizes for various stability levels
dstudy(g, stability = seq(0.8, 0.95, 0.01))
# compute both 95% and 99% confidence intervals
dstudy(g, stability = 0.9, alpha = c(0.05, 0.01) / 2)
# compute 1-tailed 95% confidence intervals
dstudy(g, alpha = 0.05)
```
Description

gstudy runs a G-study with the given data, assuming a fully crossed design (all systems evaluated on the same queries). It can be used to estimate variance components, which can further be used to run a D-study with dstudy.

Usage

gstudy(data, drop = 0)

Arguments

data A data frame or matrix with the existing effectiveness scores. Systems are columns and queries are rows.
drop The fraction of worst-performing systems to drop from the data before analysis. Defaults to 0 (include all systems).

Value

An object of class gstudy, with the following components:

n.s, n.q Number of systems and number of queries of the existing data.
var.s, var.q, var.e Variance of the system, query, and residual effects.
em.s, em.q, em.e Mean squares of the system, query and residual components.
call A list with the existing data and the percentage of systems to drop.

Author(s)

Julián Urbano

References


See Also

dstudy

Examples

g <- gstudy(adhoc3)
Maps Erh02 and Phi scores from Generalizability Theory onto traditional data-based scores like
the Kendall tau correlation, AP correlation, power, minor conflict rate and major conflict rate with
2-tailed t-tests, absolute and relative sensitivity, and rooted mean squared error.

Arguments

- **Erho2**: Vector of generalizability coefficients to map from.
- **Phi**: Vector of indices of dependability to map from.

Details

Take these mappings with a grain of salt. See figure 3 in (Urbano, 20013).

Value

A vector of data-based indicator values.

Author(s)

Julián Urbano
References


See Also
dstudy

Examples

```r
g <- gstudy(adhoc3)
d <- dstudy(g)
gt2tau(d$Erho2)
gt2rmse(d$Phi)
```

synthetic4  

*Synthetic dataset no. 4.*

Description

This is the Synthetic dataset no. 4 from Table 3.2 on page 73 of Brennan (2001), recasted as a p x i design, as required on page 182.

Usage

```r
synthetic4
```

Format

A data frame with 10 columns (systems) and 12 rows (queries).

References

Index

* datasets
  adhoc3, 2
  synthetic4, 6
adhoc3, 2
dstudy, 2, 3, 4, 6

gstudy, 2, 3, 4
gt2asens (gt2tau), 5
gt2major (gt2tau), 5
gt2minor (gt2tau), 5
gt2power (gt2tau), 5
gt2rmse (gt2tau), 5
gt2rsens (gt2tau), 5
gt2tau, 3, 5
gt2tauAP (gt2tau), 5

synthetic4, 6