Package ‘rel’

March 3, 2020

Version 1.4.2

Title Reliability Coefficients

Author Riccardo LoMartire

Maintainer Riccardo LoMartire <lomartire@kth.se>

Description Point estimates with confidence intervals for Bennett et als S, Cohen's kappa, Conger's kappa, Fleiss' kappa, Gwet's AC, intraclass correlation coefficients, Krippendorff's alpha, Scott's pi, the standard error of measurement, and weighted kappa.

Depends R (>= 2.14.0), graphics, grDevices, stats, utils

License GPL-3

NeedsCompilation no

Repository CRAN

Date/Publication 2020-03-03 08:30:02 UTC

R topics documented:

rel-package 2
bags 2
ckap 3
gac 5
icc 6
kra 8
sem 9
spi 10

Index **13**

rel-package *Reliability coefficients*

Description

`rel` provides functions for evaluating agreement between measurements: Bennett et als S (`bags`), Cohen's kappa, weighted kappa, and Conger's kappa (`ckap`), Gwet's AC1 and AC2 (`gac`), intraclass correlation coefficients (`icc`), Krippendorff's alpha (`kra`), Scott's pi and Fleiss' kappa (`spi`), and the standard error of measurement (`sem`).

bags *Bennett, Alpert, and Goldstein's S*

Description

Calculates S as an index of agreement for two observations of nominal scale data.

Usage

```
bags(data = NULL, kat = NULL, conf.level = 0.95)
```

Arguments

`data`	A matrix with n subjects and two observations (n*2 matrix)
`kat`	Number of possible categories
`conf.level`	Confidence level of the interval.

Value

`method`	Analysis name
`obs`	Number of observations
`sample`	Sample size
`est`	Point estimate
`se`	Standard error
`lb`	Lower confidence boundary
`ub`	Upper confidence boundary
`cont.table`	contingency table
`data`	analyzed data

Author(s)

Riccardo Lo Martire

References

Bennett, E. M., Alpert, R., Goldstein, A. C. (1954). Communications through limited-response questioning. Public Opinion Quarterly, 18(3), 303-308.

Examples

```
#Sample data: 200 subjects and one 5-category item.
data <- cbind(sample(1:5,200, replace=TRUE),sample(1:5,200, replace=TRUE))

#Analysis
bags(data=data, kat=5, conf.level=0.95)
```

---

ckap — *Cohen's kappa, weighted kappa, and Conger's kappa*

Description

Calculates Cohen's kappa or weighted kappa as indices of agreement for two observations of nominal or ordinal scale data, respectively, or Conger's kappa as an index of agreement for more than two observations of nominal scale data.

Usage

```
ckap(data = NULL, weight = c("unweighted", "linear", "quadratic"),
    std.err = c("Fleiss", "Cohen"), conf.level = 0.95, R = 0)
```

Arguments

`data`	A matrix with n subjects and m observations (n*m matrix).
`weight`	A character string specifying "unweighted", "linear", or "quadratic", or a numeric categories*categories matrix with custom weights (see details).
`std.err`	Standard error calculation formula specified as "Fleiss" or "Cohen" (see details).
`conf.level`	Confidence level of the interval.
`R`	Number of bootstrap replicates used to estimate the confidence interval for Conger's kappa.

Details

Cohen's kappa measures the chance-corrected agreement for two observations (Cohen, 1960 and 1968), and Conger's kappa is a generalization of Cohen's kappa for m observations (Conger, 1980). Because the maximum value for kappa commonly is restricted below 1.00 by the marginal distributions (Cohen 1960), it can also be beneficial to consider kmax when interpreting results.

By default, the standard error of Cohen's kappa is derived via Fleiss et als., corrected formula from 1969, with Cohen's original formula from 1960 optional, and the confidence interval is based on a t distribution. The confidence interval of Conger's kappa is derived via bootstrapping. Weighted kappa is based on weighted dissimilarities (diagonal = 1, off-diagonal < 1). Linear weights decrease equally with distance from the diagonal and quadratic weights decrease exponentially with distance from the diagonal. Custom weights should be specified as a categories*categories matrix with values <= 1. Incomplete cases are omitted listwise.

Value

`method`	Analysis name
`obs`	Number of observations
`sample`	Sample size
`est`	Point estimate
`se`	Standard error
`lb`	Lower confidence boundary
`ub`	Upper confidence boundary
`kmax`	The maximum value of kappa permitted by the marginal distributions
`kmax.prop`	The proportion of the kappa point estimate to the maximum possible kappa value
`cont.table`	contingency table
`data`	analyzed data

Author(s)

Riccardo Lo Martire

References

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37-46.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological Bulletin, 70(4), 213-220.

Conger, A. J. (1980). Integration and generalization of kappas for multiple raters. Psychological Bulletin, 88(2), 322-328.

Fleiss, J. L., Cohen, J., Everitt, B.S. (1969). Large sample standard errors of kappa and weighted kappa. Psychological Bulletin, 72(5), 323-327.

Examples

```
#Sample data: 200 subjects and 5 reponse categories.
data <- cbind(sample(1:5,200, replace=TRUE),sample(1:5,200, replace=TRUE))

#A numeric categories*categories matrix with custom weights
cw <- diag(ncol(matrix(0,5,5)))
cw[cw!=diag(cw)] <- runif(20,0,1)

#Cohen's kappa with Fleiss corrected standard error formula
```

```
ckap(data=data, weight="unweighted", std.err="Fleiss", conf.level = 0.95)

#Weighted kappa with linear weight
ckap(data=data, weight="linear", conf.level = 0.95)

#Weighted kappa with custom weights
ckap(data=data, weight=cw, conf.level = 0.95)
```

`gac` *Gwet's AC1 and AC2*

Description

Calculates Gwet's AC as index of agreement for two observations of nominal, ordinal, or ratio scale data.

Usage

```
gac(data = NULL, kat = NULL, weight = c("unweighted","linear","quadratic","ratio"),
    conf.level = 0.95)
```

Arguments

`data`	A matrix with n subjects and two observations (n*2 matrix)
`kat`	Number of possible categories
`weight`	A character string specifying "unweighted", "linear", "quadratic" or "ratio", or a numeric kat*kat matrix with custom weights (see details).
`conf.level`	Confidence level of the interval.

Details

Gwet's AC has the advantage of not relying on independence between observations (Gwet, 2008), making it suitable for data with dependent measurements. Weights are based on weighted dissimilarities (diagonal = 1, off-diagonal < 1). Linear weights decrease equally with distance from the diagonal and quadratic weights decrease exponentially with distance from the diagonal. Custom weights should be specified as a kat*kat matrix with values <= 1. Incomplete cases are omitted listwise and the confidence interval is based on a t distribution.

Value

`method`	Analysis name
`obs`	Number of observations
`sample`	Sample size
`est`	Point estimate
`se`	Standard error
`lb`	Lower confidence boundary

ub	Upper confidence boundary
cont.table	contingency table
data	analyzed data

Author(s)

Riccardo Lo Martire

References

Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. British Journal of Mathematical and Statistical Psychology 61, 29-48.

Examples

```
#Sample data: 200 subjects and one 5-category item.
data <- cbind(sample(1:5,200, replace=TRUE),sample(1:5,200, replace=TRUE))

#A numeric kat*kat matrix with custom weights
cw <- diag(ncol(matrix(0,5,5)))
cw[cw!=diag(cw)] <- runif(20,0,1)

#AC1
gac(data=data, kat=5, weight="unweighted", conf.level = 0.95)

#AC2 with custom weights
gac(data=data, kat=5, weight=cw, conf.level = 0.95)
```

---

icc *Intraclass correlation coefficients*

Description

Calculates intraclass correlation coefficients as a reliability measure for interval scale data.

Usage

```
icc(data = NULL, model = c("one", "two"), type = c("agreement", "consistency"),
          measure = c("single", "average"), conf.level = 0.95)
```

Arguments

data	A matrix with n subjects and m observations (n*m matrix).
model	A character string specifying "one" for one-way models or "two" for two-way models (See details).
type	A character string specifying "agreement" or "consistency" (See details).
measure	A character string specifying "singel" or "average" (See details).
conf.level	Confidence level of the interval.

Details

ICC measures the proportion of variance that is attributable to the objects of measurement (McGraw and Wong 1996). In the one-way model rows are random (i.e., columns are nested within rows), and in the two-way model both rows and columns are random (i.e., rows and columns are crossed).

Consistency considers observations relative to each other while absolute agreement considers the absolute difference of the observations (McGraw and Wong 1996). For example, ICC equals 1.00 for the paired scores (2,4), (4,6) and (6,8) for consistency, but only 0.67 for absolute agreement. In the one-way model, only absolute agreement is possible (McGraw and Wong 1996). The measure chosen should reflect the application of the tested item. The single measure is appropriate if the intention is to use the score from a single observation, while the average measure is suitable if the intention is to use an average score across a number of observations. The confidence interval is based on a F distribution. Incomplete cases are omitted listwise.

Value

`method`	Analysis name
`obs`	Number of observations
`sample`	Sample size
`est`	Point estimate
`se`	Standard error
`lb`	Lower confidence boundary
`ub`	Upper confidence boundary
`data`	analyzed data

Author(s)

Riccardo Lo Martire

References

McGraw, K. O., Wong, S. P. (1996), Forming inferences about some intraclass correlation coefficients. Psychological Methods, 1(1), 30-46.

Shrout, P. E., Fleiss, J. L. (1979), Intraclass correlation: uses in assessing rater reliability. Psychological Bulletin, 86(2), 420-428.

Examples

```
#Sample data: 200 subjects rated their weight twice.
data <- cbind(sample(50:100,200,replace=TRUE), sample(50:100,200,replace=TRUE))

#ICC based on a two-way random effects model of absolute agreement for a single observation
icc(data=data, model = "two", type = "agreement", measure ="single", conf.level=0.95)
```

kra — *Krippendorff's alpha*

Description

Calculates Krippendorff's alpha as an index of agreement for nominal, ordinal, interval, or ratio scale data.

Usage

```
kra(data = NULL, weight = c("nominal","ordinal","interval","ratio"),
    conf.level = 0.95, R = 0)
```

Arguments

`data`	A matrix with n subjects and m observations (n*m matrix).
`weight`	Data scale of ratings specified as "nominal, "ordinal", "interval", or "ratio".
`conf.level`	Confidence level of the interval.
`R`	Number of bootstrap replicates used to estimate the confidence interval.

Details

Krippendorff's alpha is a measure of observed disagreement relative to disagreement expected by chance which has the advantages of being applicable to multiple raters, different scale metrics, and incomplete data sets (Krippendorff, 2004 p.221-243). Noteworthy is that alpha by definition is zero when the observed disagreement equals the expected disagreement (i.e., when variance is absent which is the case for perfect agreement). The confidence interval of alpha is derived via bootstrapping because its distribution is unknown.

Value

`method`	Analysis name
`raters`	Number of raters
`sample`	Sample size
`na`	Percent of missing values
`est`	Point estimate
`lb`	Lower confidence boundary
`ub`	Upper confidence boundary
`cont.table`	contingency table
`data`	analyzed data

Author(s)

Riccardo Lo Martire

References

Krippendorff, K. (2004). Content Analysis: An Introduction to Its Methodology. Thousand Oaks, CA: Sage.

Examples

```
#Sample data: 4 raters, 12 subjects, and 5 ordinal scale response categories.
data <- cbind(rbind(1,2,3,3,2,1,4,1,2,NA,NA,NA),
              rbind(1,2,3,3,2,2,4,1,2,5,NA,3),
              rbind(NA,3,3,3,2,3,4,2,2,5,1,NA),
              rbind(1,2,3,3,2,4,4,1,2,5,1,NA))

#Krippendorff's alpha for ordinal scale data
kra(data = data, weight = "ordinal", conf.level = 0.95, R = 0)
```

---

sem *Standard error of measurement*

Description

Calculates the standard error of measurement

Usage

```
sem(data = NULL, type = c("mse", "sd", "cpd"), conf.level = 0.95)
```

Arguments

`data`	A matrix with n subjects and m observations (n*m matrix).
`type`	The method used to compute sem with a character string specifying "sd" for the within-subject standard deviation, "mse" for the square root of the ANOVA error variance, or "cpd" for the consecutive pairwise difference.
`conf.level`	Confidence level of the interval.

Details

"sd" and "mse" includes complete cases only and have a confidence interval based on a t distribution. "cpd" includes all cases, derives sem from the difference between adjacent trials, and has a confidence interval based on a chi squared distribution (Hopkins 2015). "cpd" is computed both overall and separately for consecutive trials, the latter allowing one to assess whether habituation decreases sem (Hopkins 2015).

Value

method	Analysis name
obs	Number of observations
sample	Sample size
na	missing values
est	Point estimate
lb	Lower confidence boundary
ub	Upper confidence boundary
est.cpd	sem for adjacent columns
data	analyzed data

Author(s)

Riccardo Lo Martire

References

Nunnally, J. C., Bernstein, I. H. (1994). Psychometric theory. New York, NY: McGraw-Hill.

Hopkins, W. G. (2015). Spreadsheets for Analysis of Validity and Reliability. Sportscience 19, 36-42.

Examples

```
#Sample data: 200 subjects rated their weight twice.
data <- cbind(sample(50:100,200,replace=TRUE), sample(50:100,200,replace=TRUE))

#Standard error of measurement
sem(data=data, type="mse", conf.level=0.95)
```

---

spi *Scott's pi and Fleiss' kappa*

Description

Calculates Scott's pi as an index of agreement for two observations of nominal or ordinal scale data, or Fleiss' kappa as an index of agreement for more than two observations of nominal scale data.

Usage

```
spi(data = NULL, weight = c("unweighted","linear","quadratic"), conf.level = 0.95)
```

Arguments

`data`	A matrix with n subjects and m observations (n*m matrix)
`weight`	A character string specifying "unweighted", "linear", or "quadratic", or a numeric categories*categories matrix with custom weights (see details).
`conf.level`	Confidence level of the interval.

Details

Scott's pi measures the chance-corrected agreement for two observations (Scott, 1955), and Fleiss' kappa is a generalization of Scott's pi for m observations (Fleiss, 1971).

The standard error for Fleiss' kappa is based on the formula from Fleiss et al., 1979. Weights are based on weighted dissimilarities (diagonal = 1, off-diagonal < 1). Linear weights decrease equally with distance from the diagonal and quadratic weights decrease exponentially with distance from the diagonal. Custom weights should be specified as a categories*categories matrix with values <= 1. Incomplete cases are omitted listwise and the confidence interval is based on a t distribution.

Value

`method`	Analysis name
`obs`	Number of observations
`sample`	Sample size
`est`	Point estimate
`se`	Standard error
`lb`	Lower confidence boundary
`ub`	Upper confidence boundary
`cont.table`	contingency table
`data`	analyzed data

Author(s)

Riccardo Lo Martire

References

Scott W. A. (1955). Reliability of content analysis: The case of nominal scale coding. Public Opinion Quarterly, 19(3), 321-325.

Fleiss J. L. (1971). Measuring nominal scale agreement among many raters. Psychological Bulletin, 76(5), 378-382.

Fleiss J. L., Nee J. C. M., Landis J. R. (1979). Large sample variance of kappa in the case of different sets of raters. Psychological Bulletin, 86(5), 974-977.

Examples

```
#Sample data: 200 subjects and one 5-category item.
data <- cbind(sample(1:5,200, replace=TRUE),sample(1:5,200, replace=TRUE))

#A numeric categories*categories matrix with custom weights
cw <- diag(ncol(matrix(0,5,5)))
cw[cw!=diag(cw)] <- runif(20,0,1)

#Scott's pi
spi(data=data, weight="unweighted", conf.level = 0.95)

#Weighted pi with custom weights
spi(data=data, weight=cw, conf.level = 0.95)
```

Index

∗Topic **package**
    rel-package, 2
∗Topic **univar**
    bags, 2
    ckap, 3
    gac, 5
    icc, 6
    kra, 8
    rel-package, 2
    sem, 9
    spi, 10

bags, *2*, 2

ckap, *2*, 3

gac, *2*, 5

icc, *2*, 6

kra, *2*, 8

rel *(*rel-package*)*, 2
rel-package, 2

sem, *2*, 9
spi, *2*, 10