Package 'irr'

October 13, 2022

Version 0.84.1

Date 2012-01-22

Title Various Coefficients of Interrater Reliability and Agreement

Author Matthias Gamer <m.gamer@uke.uni-hamburg.de>, Jim Lemon
<jim@bitwrit.com.au>, Ian Fellows <ifellows@uscd.edu> Puspendra
Singh <puspendra.pusp22@gmail.com>

Maintainer Matthias Gamer <m.gamer@uke.uni-hamburg.de>

Depends R (>= 2.10), lpSolve

Description Coefficients of Interrater Reliability and Agreement for
quantitative, ordinal and nominal data: ICC, Finn-Coefficient,
Robinson's A, Kendall's W, Cohen's Kappa, ...

License GPL (>= 2)

URL https://www.r-project.org

Repository CRAN

Date/Publication 2019-01-26 17:07:15 UTC

NeedsCompilation no

R topics documented:

---

agree	*Simple and extended percentage agreement*

Description

Computes simple and extended percentage agreement among raters.

Usage

```
agree(ratings, tolerance=0)
```

Arguments

ratings	n*m matrix or dataframe, n subjects m raters.
tolerance	number of successive rating categories that should be regarded as rater agreement (see details).

Details

Missing data are omitted in a listwise way.
Using extended percentage agreement (tolerance!=0) is only possible for numerical values. If tolerance equals 1, for example, raters differing by one scale degree are interpreted as agreeing.

Value

A list with class '"irrlist"' containing the following components:

$method	a character string describing the method applied for the computation of interrater reliability.
$subjects	the number of subjects examined.
$raters	the number of raters.
$irr.name	a character string specifying the name of the coefficient.
$value	coefficient of interrater reliability.

Author(s)

Matthias Gamer

See Also

kappa2, kappam.fleiss, kappam.light

Examples

```
data(video)
agree(video)     # Simple percentage agreement
agree(video, 1)  # Extended percentage agreement
```

---

anxiety   *Anxiety ratings by different raters*

Description

The data frame contains the anxiety ratings of 20 subjects, rated by 3 raters. Values are ranging from 1 (not anxious at all) to 6 (extremely anxious).

Usage

```
data(anxiety)
```

Format

A data frame with 20 observations on the following 3 variables.

rater1 ratings of the first rater

rater2 ratings of the second rater

rater3 ratings of the third rater

Source

artificial data

Examples

```
data(anxiety)
apply(anxiety,2,table)
```

| bhapkar | *Bhapkar coefficient of concordance between raters* |

Description

Calculates the Bhapkar coefficient of concordance for two raters.

Usage

```
bhapkar(ratings)
```

Arguments

ratings   n*2 matrix or dataframe, n subjects 2 raters.

Details

Missing data are omitted in a listwise way. The Bhapkar (1966) test is a more powerful alternative to the Stuart-Maxwell test. Both tests are asymptotically equivalent and will produce comparable chi-squared values when applied a large sample of rated objects.

Value

A list with class "irrlist" containing the following components:

$method	a character string describing the method.
$subjects	the number of data objects.
$raters	the number of raters.
$irr.name	the name of the coefficient (Chisq).
$value	the value of the coefficient.
$stat.name	the name and df of the test statistic.
$statistic	the value of the test statistic.
$p.value	the probability of the test statistic.

Author(s)

Matthias Gamer

References

Bhapkar, V.P. (1966). A note on the equivalence of two test criteria for hypotheses in categorical data. Journal of the American Statistical Association, 61, 228-235.

See Also

mcnemar.test, stuart.maxwell.mh, rater.bias

Examples

```
data(vision)
bhapkar(vision) # Original example used from Bhapkar (1966)
```

---

diagnoses	*Psychiatric diagnoses provided by different raters*

Description

Psychiatric diagnoses of n=30 patients provided by different sets of m=6 raters. Data were used by Fleiss (1971) to illustrate the computation of Kappa for m raters.

Usage

```
data(diagnoses)
```

Format

A data frame with 30 observations (psychiatric diagnoses with levels 1. Depression, 2. Personality Disorder, 3. Schizophrenia, 4. Neurosis, 5. Other) on 6 variables representing different raters.

rater1 a factor including the diagnoses of rater 1 (levels see above)

rater2 a factor including the diagnoses of rater 2 (levels see above)

rater3 a factor including the diagnoses of rater 3 (levels see above)

rater4 a factor including the diagnoses of rater 4 (levels see above)

rater5 a factor including the diagnoses of rater 5 (levels see above)

rater6 a factor including the diagnoses of rater 6 (levels see above)

Source

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

References

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

Examples

```
data(diagnoses)
table(diagnoses[,1])
```

---

finn   *Finn coefficient for oneway and twoway models*

Description

Computes the Finn coefficient as an index of the interrater reliability of quantitative data. Additionally, F-test and confidence interval are computed.

Usage

```
finn(ratings, s.levels, model = c("oneway", "twoway"))
```

Arguments

ratings	n*m matrix or dataframe, n subjects m raters.
s.levels	the number of different rating categories.
model	a character string specifying if a '"oneway"' model (default) with row effects random, or a '"twoway"' model with column and row effects random should be applied. You can specify just the initial letter.

Details

Missing data are omitted in a listwise way.

The Finn coefficient is especially useful, when variance between raters is low (i.e. agreement is high).

For the computation it could be specified if only the subjects are considered as random effects ('"oneway"' model) or if subjects and raters are randomly chosen from a bigger pool of persons ('"twoway"' model).

Value

A list with class '"irrlist"' containing the following components:

$method	a character string describing the method applied for the computation of interrater reliability.
$subjects	the number of subjects examined.
$raters	the number of raters.
$irr.name	a character string specifying the name of the coefficient.
$value	coefficient of interrater reliability.
$stat.name	a character string specifying the name and the df of the corresponding F-statistic.
$statistic	the value of the test statistic.
$p.value	the p-value for the test.

Author(s)

Matthias Gamer

References

Finn, R.H. (1970). A note on estimating the reliability of categorical data. Educational and Psychological Measurement, 30, 71-76.

See Also

[icc](#), [meancor](#), [robinson](#)

Examples

```
data(video)
finn(video, 6, model="twoway")
```

---

icc   *Intraclass correlation coefficient (ICC) for oneway and twoway models*

Description

Computes single score or average score ICCs as an index of interrater reliability of quantitative data. Additionally, F-test and confidence interval are computed.

Usage

```
icc(ratings, model = c("oneway", "twoway"),
    type = c("consistency", "agreement"),
    unit = c("single", "average"), r0 = 0, conf.level = 0.95)
```

Arguments

ratings	n*m matrix or dataframe, n subjects m raters.
model	a character string specifying if a '"oneway"' model (default) with row effects random, or a '"twoway"' model with column and row effects random should be applied. You can specify just the initial letter.
type	a character string specifying if '"consistency"' (default) or '"agreement"' between raters should be estimated. If a '"oneway"' model is used, only '"consistency"' could be computed. You can specify just the initial letter.
unit	a character string specifying the unit of analysis: Must be one of '"single"' (default) or '"average"'. You can specify just the initial letter.
r0	specification of the null hypothesis r = r0. Note that a one sided test (H1: r > r0) is performed.
conf.level	confidence level of the interval.

Details

Missing data are omitted in a listwise way.

When considering which form of ICC is appropriate for an actual set of data, one has take several decisions (Shrout & Fleiss, 1979):

1. Should only the subjects be considered as random effects ('"oneway"' model) or are subjects and raters randomly chosen from a bigger pool of persons ('"twoway"' model).

2. If differences in judges' mean ratings are of interest, interrater '"agreement"' instead of '"consistency"' should be computed.

3. If the unit of analysis is a mean of several ratings, unit should be changed to '"average"'. In most cases, however, single values (unit='"single"') are regarded.

Value

A list with class '"icclist"' containing the following components:

$subjects	the number of subjects examined.
$raters	the number of raters.
$model	a character string describing the selected model for the analysis.
$type	a character string describing the selected type of interrater reliability.
$unit	a character string describing the unit of analysis.
$icc.name	a character string specifying the name of ICC according to McGraw & Wong (1996).
$value	the intraclass correlation coefficient.
$r0	the specified null hypothesis.
$Fvalue	the value of the F-statistic.
$df1	the numerator degrees of freedom.
$df2	the denominator degrees of freedom.
$p.value	the p-value for a two-sided test.
$conf.level	the confidence level for the interval.
$lbound	the lower bound of the confidence interval.
$ubound	the upper bound of the confidence interval.

Author(s)

Matthias Gamer

References

Bartko, J.J. (1966). The intraclass correlation coefficient as a measure of reliability. Psychological Reports, 19, 3-11.

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

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

See Also

finn, meancor, robinson

Examples

```
data(anxiety)
icc(anxiety, model="twoway", type="agreement")

r1 <- round(rnorm(20, 10, 4))
r2 <- round(r1 + 10 + rnorm(20, 0, 2))
r3 <- round(r1 + 20 + rnorm(20, 0, 2))
icc(cbind(r1, r2, r3), "twoway")              # High consistency
icc(cbind(r1, r2, r3), "twoway", "agreement") # Low agreement
```

---

iota	*iota coefficient for the interrater agreement of multivariate observations*

Description

Computes iota as an index of interrater agreement of quantitative or nominal multivariate observations.

Usage

```
iota(ratings, scaledata = c("quantitative","nominal"),
     standardize = FALSE)
```

Arguments

ratings	list of n*m matrices or dataframes with one list element for each variable, n subjects m raters.
scaledata	a character string specifying if the data is '"quantitative"' (default) or '"nominal"'. If the data is organized in factors, '"nominal"' is chosen automatically. You can specify just the initial letter.
standardize	a logical indicating whether quantitative data should be z-standardized within each variable before the computation of iota.

Details

Each list element must contain observations for each rater and subject without missing values.
In case of one categorical variable (only one list element), iota reduces to the Fleiss exact kappa
coefficient, which was proposed by Conger (1980).

Value

A list with class '"irrlist"' containing the following components:

$method	a character string describing the method applied for the computation of interrater reliability.
$subjects	the number of subjects examined.
$raters	the number of raters.
$irr.name	a character string specifying the name of the coefficient.
$value	value of iota.
$detail	a character string specifying if the values were z-standardized before the computation of iota.

Author(s)

Matthias Gamer

References

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

Janson, H., & Olsson, U. (2001). A measure of agreement for interval or nominal multivariate
observations. Educational and Psychological Measurement, 61, 277-289.

See Also

icc, kappam.fleiss

Examples

```
data(diagnoses)
iota(list(diagnoses))  # produces the same result as...
kappam.fleiss(diagnoses, exact=TRUE)

# Example from Janson & Olsson (2001), Table 1
photo <- list()
photo[[1]] <- cbind(c( 71, 73, 86, 59, 71),  # weight ratings
                    c( 74, 80,101, 62, 83),
                    c( 76, 80, 93, 66, 77))
photo[[2]] <- cbind(c(166,160,187,161,172),  # height rating
                    c(171,170,174,163,182),
                    c(171,165,185,162,181))
iota(photo)
iota(photo, standardize=TRUE) # iota over standardized values
```

---

kappa2	*Cohen's Kappa and weighted Kappa for two raters*

Description

Calculates Cohen's Kappa and weighted Kappa as an index of interrater agreement between 2 raters on categorical (or ordinal) data. Own weights for the various degrees of disagreement could be specified.

Usage

```
kappa2(ratings, weight = c("unweighted", "equal", "squared"), sort.levels = FALSE)
```

Arguments

ratings	n*2 matrix or dataframe, n subjects 2 raters.
weight	either a character string specifying one predifined set of weights or a numeric vector with own weights (see details).
sort.levels	boolean value describing whether factor levels should be (re-)sorted during the calculation.

Details

Missing data are omitted in a listwise way.

During computation, ratings are converted to factors. Therefore, the categories are ordered accordingly. When ratings are numeric, a sorting of factor levels occurs automatically. Otherwise, levels are sorted when the function is called with sort.levels=TRUE.

kappa2 allows for calculating weighted Kappa coefficients. Beneath '"unweighted"' (default), predifined sets of weights are '"equal"' (all levels disagreement between raters are weighted equally) and '"squared"' (disagreements are weighted according to their squared distance from perfect agreement). The weighted Kappa coefficient with '"squared"' weights equals the product moment correlation under certain conditions. Own weights could be specified by supplying the function with a numeric vector of weights, starting from perfect agreement to worst disagreement. The length of this vector must equal the number of rating categories.

Value

A list with class '"irrlist"' containing the following components:

$method	a character string describing the method and the weights applied for the computation of weighted Kappa.
$subjects	the number of subjects examined.
$raters	the number of raters (=2).
$irr.name	a character string specifying the name of the coefficient.
$value	value of Kappa.
$stat.name	a character string specifying the name of the corresponding test statistic.

`$statistic`	the value of the test statistic.
`$p.value`	the p-value for the test.

Author(s)

Matthias Gamer

References

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

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

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

See Also

cor, kappa2, kappam.light

Examples

```
data(anxiety)
kappa2(anxiety[,1:2], "squared") # predefined set of squared weights
kappa2(anxiety[,1:2], (0:5)^2)   # same result with own set of squared weights

# own weights increasing gradually with larger distance from perfect agreement
kappa2(anxiety[,1:2], c(0,1,2,4,7,11))

data(diagnoses)
# Unweighted Kappa for categorical data without a logical order
kappa2(diagnoses[,2:3])
```

---

kappam.fleiss | *Fleiss' Kappa for m raters*

Description

Computes Fleiss' Kappa as an index of interrater agreement between m raters on categorical data. Additionally, category-wise Kappas could be computed.

Usage

```
kappam.fleiss(ratings, exact = FALSE, detail = FALSE)
```

Arguments

ratings	n*m matrix or dataframe, n subjects m raters.
exact	a logical indicating whether the exact Kappa (Conger, 1980) or the Kappa described by Fleiss (1971) should be computed.
detail	a logical indicating whether category-wise Kappas should be computed

Details

Missing data are omitted in a listwise way.

The coefficient described by Fleiss (1971) does not reduce to Cohen's Kappa (unweighted) for m=2 raters. Therefore, the exact Kappa coefficient, which is slightly higher in most cases, was proposed by Conger (1980).

The null hypothesis Kappa=0 could only be tested using Fleiss' formulation of Kappa.

Value

A list with class '"irrlist"' containing the following components:

$method	a character string describing the method applied for the computation of interrater reliability.
$subjects	the number of subjects examined.
$raters	the number of raters.
$irr.name	a character string specifying the name of the coefficient.
$value	value of Kappa.
$stat.name	a character string specifying the name of the corresponding test statistic.
$statistic	the value of the test statistic.
$p.value	the p-value for the test.
$detail	a table with category-wise kappas and the corresponding test statistics.

Author(s)

Matthias Gamer

References

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

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

Fleiss, J.L., Levin, B., & Paik, M.C. (2003). Statistical Methods for Rates and Proportions, 3rd Edition. New York: John Wiley & Sons.

See Also

kappa2, kappam.light

Examples

```
data(diagnoses)
kappam.fleiss(diagnoses)              # Fleiss' Kappa
kappam.fleiss(diagnoses, exact=TRUE)  # Exact Kappa
kappam.fleiss(diagnoses, detail=TRUE) # Fleiss' and category-wise Kappa

kappam.fleiss(diagnoses[,1:4])        # Fleiss' Kappa of raters 1 to 4
```

---

kappam.light   *Light's Kappa for m raters*

Description

Computes Light's Kappa as an index of interrater agreement between m raters on categorical data.

Usage

```
kappam.light(ratings)
```

Arguments

ratings   n*m matrix or dataframe, n subjects m raters.

Details

Missing data are omitted in a listwise way.
Light's Kappa equals the average of all possible combinations of bivariate Kappas between raters.

Value

A list with class '"irrlist"' containing the following components:

$method	a character string describing the method applied for the computation of interrater reliability.
$subjects	the number of subjects examined.
$raters	the number of raters.
$irr.name	a character string specifying the name of the coefficient.
$value	value of Kappa.
$stat.name	a character string specifying the name of the corresponding test statistic.
$statistic	the value of the test statistic.
$p.value	the p-value for the test.

Author(s)

Matthias Gamer

References

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

Light, R.J. (1971). Measures of response agreement for qualitative data: Some generalizations and alternatives. Psychological Bulletin, 76, 365-377.

See Also

kappa2, kappam.fleiss

Examples

```
data(diagnoses)
kappam.light(diagnoses)   # Light's Kappa
```

---

kendall   *Kendall's coefficient of concordance W*

Description

Computes Kendall's coefficient of concordance as an index of interrater reliability of ordinal data. The coefficient could be corrected for ties within raters.

Usage

```
kendall(ratings, correct = FALSE)
```

Arguments

ratings   n*m matrix or dataframe, n subjects m raters.

correct   a logical indicating whether the coefficient should be corrected for ties within raters.

Details

Missing data are omitted in a listwise way.
Kendall's W should be corrected for ties if raters did not use a true ranking order for the subjects.
A test for the significance of Kendall's W is only valid for large samples.

Value

A list with class '"irrlist"' containing the following components:

$method   a character string describing the method applied for the computation of interrater reliability.

$subjects   the number of subjects examined.

`$raters`	the number of raters.
`$irr.name`	a character string specifying the name of the coefficient.
`$value`	coefficient of interrater reliability.
`$stat.name`	a character string specifying the name and the df of the corresponding chi-squared test.
`$statistic`	the value of the test statistic.
`$p.value`	the p-value for the test.
`$error`	the character string of a warning message if ties were found within raters.

Author(s)

Matthias Gamer

References

Kendall, M.G. (1948). Rank correlation methods. London: Griffin.

See Also

[cor](), [meanrho]()

Examples

```
data(anxiety)
kendall(anxiety, TRUE)
```

---

kripp.alpha	*calculate Krippendorff's alpha reliability coefficient*

Description

calculates the alpha coefficient of reliability proposed by Krippendorff

Usage

```
 kripp.alpha(x, method=c("nominal","ordinal","interval","ratio"))
```

Arguments

x	classifier x object matrix of classifications or scores
method	data level of x

Value

A list with class '"irrlist"' containing the following components:

`$method`	a character string describing the method.
`$subjects`	the number of data objects.
`$raters`	the number of raters.
`$irr.name`	a character string specifying the name of the coefficient.
`$value`	value of alpha.
`$stat.name`	here "nil" as there is no test statistic.
`$statistic`	the value of the test statistic (NULL).
`$p.value`	the probability of the test statistic (NULL).
`cm`	the concordance/discordance matrix used in the calculation of alpha
`data.values`	a character vector of the unique data values
`levx`	the unique values of the ratings
`nmatchval`	the count of matches, used in calculation
`data.level`	the data level of the ratings ("nominal","ordinal", "interval","ratio")

Note

Krippendorff's alpha coefficient is particularly useful where the level of measurement of classification data is higher than nominal or ordinal.

Author(s)

Jim Lemon

References

Krippendorff, K. (1980). Content analysis: An introduction to its methodology. Beverly Hills, CA: Sage.

Examples

```
# the "C" data from Krippendorff
nmm<-matrix(c(1,1,NA,1,2,2,3,2,3,3,3,3,3,3,3,3,2,2,2,2,1,2,3,4,4,4,4,4,
1,1,2,1,2,2,2,2,NA,5,5,5,NA,NA,1,1,NA,NA,3,NA),nrow=4)
# first assume the default nominal classification
kripp.alpha(nmm)
# now use the same data with the other three methods
kripp.alpha(nmm,"ordinal")
kripp.alpha(nmm,"interval")
kripp.alpha(nmm,"ratio")
```

---

maxwell   *Maxwell's RE coefficient for binary data*

Description

Computes Maxwell's RE as an index of the interrater agreement of binary data.

Usage

```
maxwell(ratings)
```

Arguments

ratings   n*2 matrix or dataframe, n subjects 2 raters.

Details

Missing data are omitted in a listwise way.

Value

A list with class '"irrlist"' containing the following components:

$method   a character string describing the method applied for the computation of interrater
                   reliability.

$subjects   the number of subjects examined.

$raters   the number of raters (=2).

$irr.name   a character string specifying the name of the coefficient.

$value   value of RE.

Author(s)

Matthias Gamer

References

Maxwell, A.E. (1977). Coefficients of agreement between observers and their interpretation. British
Journal of Psychiatry, 130, 79-83.

See Also

[kappa2](#)

Examples

```
data(anxiety)
# Median-split to generate binary data
r1 <- ifelse(anxiety$rater1<median(anxiety$rater1),0,1)
r2 <- ifelse(anxiety$rater2<median(anxiety$rater2),0,1)
maxwell(cbind(r1,r2))
```

---

meancor	*Mean of bivariate correlations between raters*

Description

Computes the mean of bivariate Pearson's product moment correlations between raters as an index of the interrater reliability of quantitative data.

Usage

```
meancor(ratings, fisher = TRUE)
```

Arguments

ratings	n*m matrix or dataframe, n subjects m raters.
fisher	a logical indicating whether the correlation coefficients should be Fisher z-standardized before averaging.

Details

Missing data are omitted in a listwise way.

The mean of bivariate correlations should not be used as an index of interrater reliability when the variance of ratings differs between raters.

The null hypothesis r=0 could only be tested when Fisher z-standardized values are used for the averaging.

When computing Fisher z-standardized values, perfect correlations are omitted before averaging because z equals +/-Inf in that case.

Value

A list with class '"irrlist"' containing the following components:

$method	a character string describing the method applied for the computation of interrater reliability.
$subjects	the number of subjects examined.
$raters	the number of raters.
$irr.name	a character string specifying the name of the coefficient.
$value	coefficient of interrater reliability.
$stat.name	a character string specifying the name of the corresponding test statistic.

`$statistic`	the value of the test statistic.
`$p.value`	the p-value for the test.
`$error`	a character string specifying whether correlations were dropped before the computation of the Fisher z-standardized average.

Author(s)

Matthias Gamer

See Also

[cor](cor)

Examples

```
data(anxiety)
meancor(anxiety)
```

---

meanrho	*Mean of bivariate rank correlations between raters*

Description

Computes the mean of bivariate Spearman's rho rank correlations between raters as an index of the interrater reliability of ordinal data.

Usage

```
meanrho(ratings, fisher = TRUE)
```

Arguments

ratings	n*m matrix or dataframe, n subjects m raters.
fisher	a logical indicating whether the correlation coefficients should be Fisher z-standardized before averaging.

Details

Missing data are omitted in a listwise way.
The mean of bivariate rank correlations should not be used as an index of interrater reliability when ties within raters occur.
The null hypothesis r=0 could only be tested when Fisher z-standardized values are used for the averaging.
When computing Fisher z-standardized values, perfect correlations are omitted before averaging because z equals +/-Inf in that case.

Value

A list with class '"irrlist"' containing the following components:

$method	a character string describing the method applied for the computation of interrater reliability.
$subjects	the number of subjects examined.
$raters	the number of raters.
$irr.name	a character string specifying the name of the coefficient.
$value	coefficient of interrater reliability.
$stat.name	a character string specifying the name of the corresponding test statistic.
$statistic	the value of the test statistic.
$p.value	the p-value for the test.
$error	a character specifying whether correlations were dropped before the computation of the Fisher z-standardized average. Additionally, a warning message is created if ties were found within raters.

Author(s)

Matthias Gamer

See Also

cor, kendall

Examples

```
data(anxiety)
meanrho(anxiety, TRUE)
```

N.cohen.kappa   *Sample Size Calculation for Cohen's Kappa Statistic*

Description

This function is a sample size estimator for the Cohen's Kappa statistic for a binary outcome. Note that any value of "kappa under null" in the interval [0,1] is acceptable (i.e. k0=0 is a valid null hypothesis).

Usage

```
 N.cohen.kappa(rate1, rate2, k1, k0, alpha=0.05,
               power=0.8, twosided=FALSE)
```

Arguments

rate1	the probability that the first rater will record a positive diagnosis
rate2	the probability that the second rater will record a positive diagnosis
k1	the true Cohen's Kappa statistic
k0	the value of kappa under the null hypothesis
alpha	type I error of test
power	the desired power to detect the difference between true kappa and hypothetical kappa
twosided	TRUE if test is two-sided

Value

returns required sample size

Author(s)

Ian Fellows

References

Cantor, A. B. (1996) Sample-size calculation for Cohen's kappa. Psychological Methods, 1, 150-153.

See Also

kappa2

Examples

```
# Testing H0: kappa = 0.7 vs. HA: kappa > 0.7 given that
# kappa = 0.85 and both raters classify 50% of subjects as positive.
N.cohen.kappa(0.5, 0.5, 0.7, 0.85)
```

N2.cohen.kappa   *Sample Size Calculation for Cohen's Kappa Statistic with more than one category*

Description

This function calculates the required sample size for the Cohen's Kappa statistic when two raters have the same marginal. Note that any value of "kappa under null" in the interval [-1,1] is acceptable (i.e. k0=0 is a valid null hypothesis).

Usage

```
N2.cohen.kappa(mrg, k1, k0, alpha=0.05, power=0.8, twosided=FALSE)
```

Arguments

`mrg`	a vector of marginal probabilities given by raters
`k1`	the true Cohen's Kappa statistic
`k0`	the value of kappa under the null hypothesis
`alpha`	type I error of test
`power`	the desired power to detect the difference between true kappa and hypothetical kappa
`twosided`	TRUE if test is two-sided

Value

Returns required sample size.

Author(s)

Puspendra Singh and Jim Lemon

References

Flack, V.F., Afifi, A.A., Lachenbruch, P.A., & Schouten, H.J.A. (1988). Sample size determinations for the two rater kappa statistic. Psychometrika, 53, 321-325.

See Also

N.cohen.kappa, kappa2

Examples

```
require(lpSolve)
# Testing H0: kappa = 0.4 vs. HA: kappa > 0.4 (=0.6) given that
# Marginal Probabilities by two raters are (0.2, 0.25, 0.55).
#
# one sided test with 80% power:
N2.cohen.kappa(c(0.2, 0.25, 0.55), k1=0.6, k0=0.4)
# one sided test with 90% power:
N2.cohen.kappa(c(0.2, 0.25, 0.55), k1=0.6, k0=0.4, power=0.9)

# Marginal Probabilities by two raters are (0.2, 0.05, 0.2, 0.05, 0.2, 0.3)
# Testing H0: kappa = 0.1 vs. HA: kappa > 0.1 (=0.5) given that
#
# one sided test with 80% power:
N2.cohen.kappa(c(0.2, 0.05, 0.2, 0.05, 0.2, 0.3), k1=0.5, k0=0.1)
```

---

print.icclist *Default printing function for ICC results*

Description

Prints the results of the ICC computation.

Usage

```
## S3 method for class 'icclist'
print(x, ...)
```

Arguments

x   a list with class '"icclist"' containing the results of the ICC computation.

...   further arguments passed to or from other methods.

Details

'"print.icclist"' is only a printing function and is usually not called directly.

Author(s)

Matthias Gamer

See Also

[icc](#)

Examples

```
data(anxiety)
# "print.icclist" is the default printing function of "icc"
icc(anxiety, model="twoway", type="agreement")
```

---

print.irrlist *Default printing function for various coefficients of interrater reliability*

Description

Prints the results of various functions computing coefficients of interrater reliability.

Usage

```
## S3 method for class 'irrlist'
print(x, ...)
```

Arguments

x	a list with class '"irrlist"' containing the results of the interrater reliability computation.
...	further arguments passed to or from other methods.

Details

'"print.irrlist"' is only a printing function and is usually not called directly.

Author(s)

Matthias Gamer

See Also

[bhapkar](), [finn](), [iota](), [kappa2](), [kappam.fleiss](), [kappam.light](), [kripp.alpha](), [kendall](), [maxwell](), [meancor](), [meanrho](), [rater.bias](), [robinson](), [stuart.maxwell]()

Examples

```
data(anxiety)
# "print.irrlist" is the default printing method of various functions, e.g.
finn(anxiety, 6)
meancor(anxiety)
```

---

rater.bias   *Coefficient of rater bias*

Description

Calculates a coefficient of systematic bias between two raters.

Usage

```
rater.bias(x)
```

Arguments

x	c x c classification matrix or 2 x n or n x 2 matrix of classification scores into c categories.

Details

rater.bias calculates a reliability coefficient for two raters classifying n objects into any number of categories. It will accept either a c x c classification matrix of counts of objects falling into c categories or a 2 x n or n x 2 matrix of classification scores.
The function returns the absolute value of the triangular off-diagnonal sum ratio of the cxc classification table and the corresponding test statistic. A systematic bias between raters can be assumed when the ratio substantially deviates from 0.5 while yielding a significant Chi-squared statistic.

Value

`method`	Name of the method
`subjects`	Number of subjects
`raters`	Number of raters (2)
`irr.name`	Name of the coefficient: ratio of triangular off-diagnonal sums
`value`	Value of the coefficient
`stat.name`	Name of the test statistic
`statistic`	Value of the test statistic
`p.value`	the probability of the df 1 Chi-square variable

Author(s)

Jim Lemon

References

Bishop Y.M.M., Fienberg S.E., & Holland P.W. (1978). Discrete multivariate analysis: theory and practice. Cambridge, Massachusetts: MIT Press.

See Also

mcnemar.test

Examples

```
# fake a 2xn matrix of three way classification scores
ratings <- matrix(sample(1:3,60,TRUE), nrow=2)
rater.bias(ratings)

# Example from Bishop, Fienberg & Holland (1978), Table 8.2-1
data(vision)
rater.bias(vision)
```

---

relInterIntra	*Inter- and intra-rater reliability*

Description

'relInterIntra' calculates inter- and intra-rater reliability coefficients.

Usage

```
relInterIntra(x, nrater=1, raterLabels=NULL, rho0inter=0.6,
              rho0intra=0.8, conf.level=.95)
```

Arguments

x	Data frame or matrix of rater by object scores
nrater	Number of raters
raterLabels	Labels for the raters or methods
rho0inter	Null hypothesis value for the inter-rater reliability coefficient
rho0intra	Null hypothesis value for the intra-rater reliability coefficient
conf.level	Confidence level for the one-sided confidence interval reported

Value

nil

Author(s)

Tore Wentzel-Larsen

References

Eliasziw, M., Young, S.L., Woodbury, M.G., & Fryday-Field, K. (1994). Statistical methodology for the concurrent assessment of interrater and intrarater reliability: Using goniometric measurements as an example. Physical Therapy, 74, 777-788.

Examples

```
# testing code for the Goniometer data from the article:
 table4<-matrix(c(
  -2,16,5,11,7,-7,18,4,0,0,-3,3,7,-6,1,-13,2,4,-10,8,7,-3,-5,5,0,7,-8,1,-3,
  0,16,6,10,8,-8,19,5,-3,0,-2,-1,9,-7,1,-14,1,4,-9,9,6,-2,-5,5,-1,6,-8,1,-3,
  1,15,6,10,6,-8,19,5,-2,-2,-2,1,9,-6,0,-14,0,3,-10,8,7,-4,-7,5,-1,6,-8,2,-3,
  2,12,4,9,5,-9,17,5,-7,1,-4,-1,4,-8,-2,-12,-1,7,-10,2,8,-5,-6,3,-4,4,-10,1,-5,
  1,14,4,7,6,-10,17,5,-6,2,-3,-2,4,-10,-2,-12,0,6,-11,8,7,-5,-8,4,-3,4,-11,-1,-4,
  1,13,4,8,6,-9,17,5,-5,1,-3,1,2,-9,-3,-12,0,4,-10,8,7,-5,-7,4,-4,4,-10,0,-5
  ),ncol=6)
 relInterIntra(x=table4,nrater=2,raterLabels=c('universal','Lamoreux'))
```

---

robinson	*Robinson's A*

Description

Computes Robinson's A as an index of the interrater reliability of quantitative data.

Usage

```
robinson(ratings)
```

Arguments

ratings   n*m matrix or dataframe, n subjects m raters.

Details

Missing data are omitted in a listwise way.

Value

A list with class '"irrlist"' containing the following components:

$method   a character string describing the method applied for the computation of interrater reliability.

$subjects   the number of subjects examined.

$raters   the number of raters.

$irr.name   a character string specifying the name of the coefficient.

$value   coefficient of interrater reliability.

Author(s)

Matthias Gamer

References

Robinson, W.S. (1957). The statistical measurement of agreement. American Sociological Review, 22, 17-25.

See Also

[finn](), [icc](), [meancor]()

Examples

```
data(anxiety)
robinson(anxiety)
```

---

stuart.maxwell.mh   *Stuart-Maxwell coefficient of concordance for two raters*

Description

Calculates the Stuart-Maxwell coefficient of concordance for two raters.

Usage

```
stuart.maxwell.mh(x)
```

Arguments

x   c x c classification matrix or matrix of classification scores into c categories.

Details

stuart.maxwell.mh calculates a reliability coefficient for two raters classifying n objects into any number of categories. It will accept either a c x c classification matrix of counts of objects falling into c categories or a c x n or n x c matrix of classification scores.

Value

A list with class "irrlist" containing the following components:

$method	a character string describing the method.
$subjects	the number of data objects.
$raters	the number of raters.
$irr.name	the name of the coefficient (Chisq).
$value	the value of the coefficient.
$stat.name	the name and df of the test statistic.
$statistic	the value of the test statistic.
$p.value	the probability of the test statistic.

Author(s)

Jim Lemon

References

Stuart, A.A. (1955). A test for homogeneity of the marginal distributions in a two-way classification. Biometrika, 42, 412-416.

Maxwell, A.E. (1970) Comparing the classification of subjects by two independent judges. British Journal of Psychiatry, 116, 651-655.

See Also

bhapkar, rater.bias

Examples

```
# fake a 2xn matrix of three way classification scores
ratings<-matrix(sample(1:3,60,TRUE), nrow=2)
stuart.maxwell.mh(ratings)

# Example used from Stuart (1955)
data(vision)
stuart.maxwell.mh(vision)
```

---

video   *Different raters judging the credibility of videotaped testimonies*

Description

The data frame contains the credibility ratings of 20 subjects, rated by 4 raters. Judgements could vary from 1 (not credible) to 6 (highly credible). Variance between and within raters is low.

Usage

```
data(video)
```

Format

A data frame with 20 observations on the following 4 variables.

rater1 ratings of rater 1

rater2 ratings of rater 2

rater3 ratings of rater 3

rater4 ratings of rater 4

Source

artificial data

Examples

```
data(video)
apply(video,2,table)
```

---

vision   *Eye-testing case records*

Description

Case records of the eye-testing of N=7477 female employees in Royal Ordnance factories between 1943 and 1946. Data were primarily used by Stuart (1953) to illustrate the the estimation and comparison of strengths of association in contingency tables.

Usage

```
data(anxiety)
```

Format

A data frame with 7477 observations (eye testing results with levels 1st grade, 2nd grade, 3rd grade, 4th Grade) on the following 2 variables.

r.eye unaided distance vision performance of the right eye

l.eye unaided distance vision performance of the left eye

Source

Stuart, A. (1953). The Estimation and Comparison of Strengths of Association in Contingency Tables. Biometrika, 40, 105-110.

References

Stuart, A. (1953). The Estimation and Comparison of Strengths of Association in Contingency Tables. Biometrika, 40, 105-110.

Examples

```
data(vision)
table(vision$r.eye, vision$l.eye)
```

Index