Package: contentValidity (via r-universe)

October 9, 2026

Type Package

Title Content Validity Indices for Instrument Development

Version 0.2.0

Description Computes content validity indices commonly used in instrument development and questionnaire validation, including the Item-level Content Validity Index (I-CVI), Scale-level Content Validity Index (S-CVI), modified kappa adjusted for chance agreement, Aiken's V, and Lawshe's Content Validity Ratio (CVR). Methods follow Lynn (1986) <doi:10.1097/00006199-198611000-00017>, Polit and Beck (2006) <doi:10.1002/nur.20147>, Aiken (1985) <doi:10.1177/0013164485451012>, and Lawshe (1975) <doi:10.1111/j.1744-6570.1975.tb01393.x>.

License MIT + file LICENSE

Encoding UTF-8

Roxygen list(markdown = TRUE)

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

VignetteBuilder knitr

Config/testthat/edition 3

URL https://rafhq1403.github.io/contentValidity/, https://github.com/Rafhq1403/contentValidity

BugReports https://github.com/Rafhq1403/contentValidity/issues

Config/roxygen2/version 8.0.0

Depends R (>= 3.5)

Imports stats, graphics, grDevices

LazyData true

Repository https://rafhq1403.r-universe.dev

Date/Publication 2026-09-09 20:53:05 UTC

RemoteUrl https://github.com/rafhq1403/contentvalidity

RemoteRef HEAD

RemoteSha 5873ec163e0cd92dba68c92b27af5175c8884776

Contents

aiken_v 2
apa_table 4
content_validity 6
cv_sample_size_icvi 8
cvi_example 10
cvr 11
cvr_critical 13
gwet_ac1 14
gwet_ac2 17
icvi 20
mod_kappa 22
plot.content_validity 24
print.content_validity 26
scvi_ave 27
scvi_ua 28

Index **31**

`aiken_v` *Aiken's V coefficient of content validity*

Description

Computes Aiken's V (Aiken, 1985), an index of content validity that uses the full rating scale rather than dichotomizing responses as in I-CVI. Aiken's V ranges from 0 to 1, where 1 indicates all experts gave the maximum rating and 0 indicates all gave the minimum.

Usage

```
aiken_v(
  ratings,
  lo = 1,
  hi = 4,
  na.rm = FALSE,
  ci = FALSE,
  n_boot = 2000,
  ci_method = c("percentile", "bca"),
  conf_level = 0.95,
  seed = NULL
)
```

Arguments

`ratings`	A numeric matrix or data frame of expert ratings (rows = experts, columns = items). A numeric vector is also accepted, treated as a single item.
`lo`	Numeric. Minimum possible rating on the scale. Default 1.

hi	Numeric. Maximum possible rating on the scale. Default 4.
na.rm	Logical. If TRUE, missing ratings are excluded. Defaults to FALSE.
ci	Logical. If TRUE, returns a data frame with bootstrap confidence intervals alongside the point estimate. Defaults to FALSE (returns a numeric vector, identical to the package's pre-0.2.0 behaviour).
n_boot	Integer. Number of bootstrap replicates when ci = TRUE. Defaults to 2000 (Davison & Hinkley, 1997; Hesterberg, 2015).
ci_method	Character. One of "percentile" (default; Efron & Tibshirani, 1993) or "bca" (bias-corrected and accelerated; DiCiccio & Efron, 1996).
conf_level	Numeric. Confidence level between 0 and 1. Defaults to 0.95.
seed	Integer or NULL. If supplied, passed to set.seed() for reproducible bootstrap samples. Defaults to NULL.

Details

Optional bootstrap confidence intervals are available via ci = TRUE. Resampling is performed at the expert (row) level, matching the standard inferential frame for inter-rater reliability analyses (Gwet, 2014).

Aiken's V is calculated as:

$$V = (\bar{X} - lo)/(hi - lo)$$

where $\bar{X}$ is the mean expert rating across raters, and lo and hi are the minimum and maximum possible scale values, respectively.

A common cutoff is V >= 0.70 for adequate content validity, though stricter thresholds are sometimes applied depending on panel size and research context. Unlike I-CVI, Aiken's V uses the full rating scale, so a rating of 4 contributes more than a rating of 3 (rather than both being counted equally as "relevant").

Value

When ci = FALSE (default), a named numeric vector of V values, one per item (or a single numeric value if ratings is a vector). When ci = TRUE, a data frame with one row per item and columns item, aiken_v, ci_lower, ci_upper, ci_method, conf_level, n_boot.

References

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131-142. doi:10.1177/0013164485451012

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge University Press. doi:10.1017/CBO9780511802843

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, 11(3), 189-228. doi:10.1214/ss/1032280214

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman and Hall. doi:10.1201/9780429246593

Gwet, K. L. (2014). *Handbook of inter-rater reliability* (4th ed.). Advanced Analytics, LLC.

Hesterberg, T. C. (2015). What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum. *The American Statistician*, 69(4), 371-386. doi:10.1080/00031305.2015.1089789

See Also

`icvi()`

Examples

```
ratings <- matrix(
  c(4, 4, 3, 4, 4,
    3, 4, 4, 4, 3,
    2, 3, 3, 4, 3,
    1, 2, 3, 2, 3),
  nrow = 5,
  dimnames = list(NULL, paste0("item", 1:4))
)
aiken_v(ratings)

# 5-point scale
aiken_v(c(5, 4, 5, 5, 4), lo = 1, hi = 5)

# With bootstrap confidence intervals (new in v0.2.0)
aiken_v(ratings, ci = TRUE, n_boot = 1000, seed = 1)
```

---

apa_table *APA-style content validity table*

Description

Generates a publication-ready content validity table following APA conventions, suitable for inclusion in journal manuscripts, theses, and technical reports. Returns a clean data frame by default, with optional rendering to markdown, HTML, or LaTeX via `knitr::kable()`.

Usage

```
apa_table(x, ...)

## S3 method for class 'content_validity'
apa_table(
  x,
  format = c("data.frame", "markdown", "html", "latex", "pipe"),
  digits = 2,
  interpretation = TRUE,
  interpretation_index = c("mod_kappa", "gwet_ac1", "gwet_ac2", "icvi"),
  caption = NULL,
  ...
)
```

Arguments

x	An object to format. Currently supports objects of class "content_validity" returned by content_validity().
...	Further arguments passed to methods.
format	Output format. One of "data.frame" (default), "markdown", "html", "latex", or "pipe". All formats other than "data.frame" require the knitr package.
digits	Integer. Number of decimal places for numeric values. Default 2 (APA convention for proportions and correlations).
interpretation	Logical. Whether to include an interpretation column. Default TRUE. The cutoffs depend on interpretation_index.
interpretation_index	Character. Which index drives the interpretation column. One of "mod_kappa" (default; Cicchetti & Sparrow, 1981; Polit, Beck, & Owen, 2007), "gwet_ac1" (Altman, 1991), "gwet_ac2" (Altman, 1991), or "icvi" (Polit & Beck, 2006). The resulting column is named accordingly (e.g., "Kappa Interpretation", "AC1 Interpretation") so that the labels are not confused with the other columns in the table.
caption	Optional character string. The caption to use when format is not "data.frame". If NULL (default), a standard caption is generated that reports the scale-level indices.

Details

Item-level interpretation labels follow the modified-kappa cutoffs of Cicchetti and Sparrow (1981), as adopted by Polit, Beck, and Owen (2007):

- Excellent: kappa* > 0.74
- Good: kappa* 0.60 to 0.74
- Fair: kappa* 0.40 to 0.59
- Poor: kappa* < 0.40

Scale-level indices are reported in the caption rather than the table body, matching the typical layout used in nursing, education, and health-sciences journals.

Value

A data frame (when format = "data.frame") or a character string suitable for inclusion in an R Markdown document (other formats).

References

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, 86(2), 127-137.

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*, 30(4), 459-467. doi:10.1002/nur.20199

Examples

```
data(cvi_example)
result <- content_validity(cvi_example)

# Default: a clean data frame
apa_table(result)

# Markdown for R Markdown documents
if (requireNamespace("knitr", quietly = TRUE)) {
  apa_table(result, format = "markdown")
}
```

`content_validity` *Comprehensive content validity analysis*

Description

Runs the standard relevance-scale content validity indices on a single ratings matrix and returns a tidy summary. Computes Item-level CVI, modified kappa, Aiken's V, Gwet's AC1, and Gwet's AC2 at the item level; S-CVI/Ave, S-CVI/UA, mean modified kappa, mean AC1, and mean AC2 at the scale level. New AC1 and AC2 columns added in v0.2.0.

Usage

```
content_validity(
  ratings,
  relevant_threshold = 3,
  lo = 1,
  hi = 4,
  categories = NULL,
  ac2_weights = "quadratic",
  subscale = NULL,
  na.rm = FALSE
)
```

Arguments

`ratings`	A numeric matrix or data frame of expert ratings (rows = experts, columns = items) on a relevance scale.
`relevant_threshold`	Integer. Minimum rating considered "relevant". Passed to `icvi()`, `scvi_ave()`, `scvi_ua()`, `mod_kappa()`, and `gwet_ac1()`. Defaults to 3.
`lo, hi`	Numeric. Minimum and maximum possible rating values on the scale; passed to `aiken_v()`. Defaults to 1 and 4.
`categories`	Numeric vector of all possible rating values, used by `gwet_ac2()`. Defaults to `seq(lo, hi)`, which is correct for the typical 4-point relevance scale.

ac2_weights	Weighting scheme passed to gwet_ac2(). One of "quadratic" (default), "linear", "identity", or a custom square matrix.
subscale	Optional character or factor vector of length ncol(ratings) assigning each item to a subscale (factor / domain). When supplied, the scale-level indices are computed both overall and per-subscale, and the result carries a $subscales data frame. Useful for multi-dimensional instruments where different items measure different constructs. Defaults to NULL (overall only).
na.rm	Logical. Passed to all underlying functions. Defaults to FALSE.

Details

Lawshe's CVR is not included in this wrapper because it uses a different rating convention (essential / useful but not essential / not necessary). For CVR analyses, use cvr() and cvr_critical() directly.

Value

An object of class "content_validity": a list containing

- items: a data frame with one row per item and columns item, icvi, mod_kappa, aiken_v, gwet_ac1, gwet_ac2.
- scale: a named numeric vector with scvi_ave, scvi_ua, mean_kappa, mean_ac1, mean_ac2.
- n_experts: integer, number of experts (rows).
- n_items: integer, number of items (columns).

See Also

icvi(), scvi_ave(), scvi_ua(), mod_kappa(), aiken_v(), gwet_ac1(), gwet_ac2(), cvr()

Examples

```
ratings <- matrix(
  c(4, 4, 3, 4, 4,
    3, 4, 4, 4, 3,
    2, 3, 3, 4, 3,
    1, 2, 3, 2, 3),
  nrow = 5,
  dimnames = list(NULL, paste0("item", 1:4))
)
result <- content_validity(ratings)
result
result$items
result$scale
```

`cv_sample_size_icvi` *Sample-size planning for content-validity studies*

Description

Computes the minimum number of expert raters required to estimate an Item-level Content Validity Index (I-CVI) within a specified confidence-interval half-width at a chosen confidence level. Two methods are supported:

Usage

```
cv_sample_size_icvi(
  expected,
  half_width,
  conf_level = 0.95,
  method = c("wald", "wilson"),
  max_n = 1000
)
```

Arguments

`expected`	Numeric in `[0, 1]`. Anticipated I-CVI value. Common values are 0.80–0.95 for items that pass review.
`half_width`	Numeric in `(0, 1)`. Desired half-width of the confidence interval. Smaller half-widths require more experts. Typical choices are 0.05–0.15.
`conf_level`	Numeric in `(0, 1)`. Confidence level. Default 0.95.
`method`	One of `"wald"` (default) or `"wilson"`.
`max_n`	Upper bound on the bisection search for the Wilson method. Defaults to 1000. If the required sample size exceeds this, the function returns `NA` with a warning.

Details

- `"wald"` (default): the closed-form normal approximation. Fast and widely used in introductory sample-size formulas. Slightly anti-conservative for I-CVI values near 0 or 1.
- `"wilson"`: the Wilson score interval (Wilson, 1927), solved numerically via `stats::uniroot()`. More accurate for proportions near 0 or 1, which is the common case in content-validity work where I-CVI is typically high (e.g., 0.80–0.95). Recommended by Newcombe (1998) and Agresti & Coull (1998) for proportion CIs in small-to-moderate samples.

The result fills a documented gap in the content-validity literature. Lynn (1986) and Polit & Beck (2006) provide rule-of-thumb recommendations (typically 5–10 experts) without statistical justification; this function gives a precision-based answer suitable for justification in study protocols and grant applications.

Wald formula:

$$n = \lceil z^2 \pi(1-\pi)/w^2 \rceil$$

where $z = \Phi^{-1}(1 - \alpha/2)$, π is the expected I-CVI, and w is the target half-width.

Wilson formula: The Wilson score interval has half-width:

$$w(n) = z\sqrt{\pi(1-\pi)/n + z^2/(4n^2)}/(1 + z^2/n)$$

which is decreasing in n. The function uses `stats::uniroot()` to find the smallest n such that $w(n) \leq w_{target}$.

At $\pi = 0.85$, $w = 0.10$, $1 - \alpha = 0.95$:

- Wald gives n = ceiling(1.96^2 * 0.85 * 0.15 / 0.10^2) = 49
- Wilson gives n = 49 (essentially identical in the central range)

At $\pi = 0.95$, $w = 0.05$:

- Wald gives n = 73
- Wilson gives n = 83 (more conservative near the boundary)

For typical content-validity targets (e.g., expected I-CVI 0.85, half-width 0.15), both methods recommend roughly 19–22 experts, well above Lynn's (1986) rule-of-thumb minimum of 6 – a useful caveat to flag in study design and grant applications.

Value

An integer: the minimum number of experts required.

References

Agresti, A., & Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, 52(2), 119-126. doi:10.1080/00031305.1998.10480550

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382-385. doi:10.1097/00006199198611000000017

Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Statistics in Medicine*, 17(8), 857-872. doi:10.1002/(SICI)10970258(19980430)17:8<857::AID-SIM777>3.0.CO;2E

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29(5), 489-497. doi:10.1002/nur.20147

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158), 209-212. doi:10.1080/01621459.1927.10502953

See Also

`icvi()`

Examples

```
# Common scenario: anticipated I-CVI = 0.85, want half-width <= 0.10
cv_sample_size_icvi(expected = 0.85, half_width = 0.10)

# More precision (half-width <= 0.05) needs more experts
cv_sample_size_icvi(expected = 0.85, half_width = 0.05)

# Wilson method is more accurate near the upper bound
cv_sample_size_icvi(expected = 0.95, half_width = 0.05,
                    method = "wilson")

# Sensitivity table over a range of expected I-CVIs
sapply(seq(0.70, 0.95, by = 0.05), function(p) {
  cv_sample_size_icvi(expected = p, half_width = 0.10)
})
```

`cvi_example` *Example expert ratings for content validity analysis*

Description

A simulated dataset illustrating typical expert ratings during the content validation of a 10-item depression screening instrument. Six expert clinicians rate each item's relevance on a 4-point scale.

Usage

```
cvi_example
```

Format

A 6 by 10 numeric matrix with rows representing expert raters (`expert1` through `expert6`) and columns representing candidate items (`item1` through `item10`). Values are on a 4-point relevance scale:

- 1: not relevant
- 2: somewhat relevant (item needs major revision)
- 3: quite relevant (item needs minor revision)
- 4: highly relevant

Details

The pattern of ratings is realistic: some items achieve universal agreement, most show strong but imperfect agreement, and a couple of items would be flagged for revision based on standard CVI cutoffs (e.g., items 5 and 9 in this example).

Source

Simulated for demonstration; not based on real expert ratings.

Examples

```
data(cvi_example)
icvi(cvi_example)
content_validity(cvi_example)
```

`cvr` *Lawshe's Content Validity Ratio (CVR)*

Description

Computes Lawshe's (1975) Content Validity Ratio for one or more items rated by an expert panel. Each expert classifies an item as "essential", "useful but not essential", or "not necessary"; CVR captures the proportion of experts endorsing "essential" relative to chance.

Usage

```
cvr(
  ratings,
  essential = 1,
  na.rm = FALSE,
  ci = FALSE,
  n_boot = 2000,
  ci_method = c("percentile", "bca"),
  conf_level = 0.95,
  seed = NULL
)
```

Arguments

`ratings`	A numeric matrix or data frame of expert ratings (rows = experts, columns = items). A numeric vector is also accepted, treated as a single item.
`essential`	Numeric vector. Rating value(s) that indicate an expert classified the item as "essential". Defaults to 1, matching Lawshe's (1975) original 3-point scale where 1 = essential, 2 = useful but not essential, 3 = not necessary. Pass a vector if multiple values count as essential.
`na.rm`	Logical. If `TRUE`, missing ratings are excluded when counting experts. Defaults to `FALSE`.
`ci`	Logical. If `TRUE`, returns a data frame with bootstrap confidence intervals alongside the point estimate. Defaults to `FALSE` (returns a numeric vector, identical to the package's pre-0.2.0 behaviour).
`n_boot`	Integer. Number of bootstrap replicates when `ci = TRUE`. Defaults to 2000 (Davison & Hinkley, 1997; Hesterberg, 2015).
`ci_method`	Character. One of `"percentile"` (default; Efron & Tibshirani, 1993) or `"bca"` (bias-corrected and accelerated; DiCiccio & Efron, 1996).
`conf_level`	Numeric. Confidence level between 0 and 1. Defaults to 0.95.
`seed`	Integer or `NULL`. If supplied, passed to `set.seed()` for reproducible bootstrap samples. Defaults to `NULL`.

Details

The formula is:

$$CVR = (n_e - N/2)/(N/2)$$

where n_e is the number of experts rating the item as essential and N is the total number of experts.

Use `cvr_critical()` to obtain the minimum CVR considered statistically significant for a given panel size, following the corrected critical values of Wilson, Pan, and Schumsky (2012).

Value

A named numeric vector of CVR values per item, ranging from -1 to +1. If `ratings` is a vector, returns a single numeric value.

References

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563-575. doi:10.1111/j.17446570.1975.tb01393.x

Wilson, F. R., Pan, W., & Schumsky, D. A. (2012). Recalculation of the critical values for Lawshe's content validity ratio. *Measurement and Evaluation in Counseling and Development*, 45(3), 197-210. doi:10.1177/0748175612440286

See Also

`cvr_critical()`

Examples

```
# 10 experts rating 3 items on Lawshe's 3-point scale
# (1 = essential, 2 = useful, 3 = not necessary)
ratings <- matrix(
  c(1, 1, 1, 1, 1, 1, 1, 1, 2, 2,    # 8 of 10 essential
    1, 1, 1, 2, 2, 2, 2, 3, 3, 3,    # 3 of 10 essential
    1, 1, 1, 1, 1, 1, 1, 1, 1, 1),   # 10 of 10 essential
  nrow = 10,
  dimnames = list(NULL, paste0("item", 1:3))
)
cvr(ratings)

# Compare to the critical value for N = 10
cvr_critical(10)

# With bootstrap confidence intervals
cvr(ratings, ci = TRUE, n_boot = 1000, seed = 1)
```

`cvr_critical` *Critical CVR value for a given panel size*

Description

Returns the minimum Content Validity Ratio considered statistically significant for a panel of N experts at the specified alpha level. The calculation uses the exact binomial distribution under the null hypothesis that each expert independently rates "essential" with probability 0.5, following the corrected approach of Wilson, Pan, and Schumsky (2012).

Usage

```
cvr_critical(n_experts, alpha = 0.05)
```

Arguments

`n_experts`	Positive integer. Number of experts on the panel.
`alpha`	Numeric. One-tailed significance level. Defaults to 0.05.

Details

The critical value is determined as the smallest k such that $P(X \geq k) \leq \alpha$ when $X \sim Binomial(N, 0.5)$, then transformed to the CVR scale via $CVR_{crit} = (k - N/2)/(N/2)$.

Wilson, Pan, and Schumsky (2012) demonstrated that Lawshe's (1975) original critical-value table contained errors, especially for small panels. The exact binomial computation used here is their recommended replacement.

Value

Numeric. The critical CVR value. CVR values at or above this threshold are statistically significant. Returns `NA_real_` if no CVR value can reach significance at the specified alpha (which can happen for very small panels with stringent alpha).

References

Wilson, F. R., Pan, W., & Schumsky, D. A. (2012). Recalculation of the critical values for Lawshe's content validity ratio. *Measurement and Evaluation in Counseling and Development*, 45(3), 197-210. doi:10.1177/0748175612440286

See Also

`cvr()`

Examples

```
cvr_critical(10)         # 0.80 -- need 9 of 10 experts to call it essential
cvr_critical(20)         # 0.50
cvr_critical(40)         # 0.25
cvr_critical(10, alpha = 0.01)
```

---

gwet_ac1 *Gwet's AC1 - chance-corrected agreement*

Description

Computes Gwet's AC1 coefficient (Gwet, 2008) for each item rated by an expert panel on a relevance scale. AC1 is a chance-corrected agreement index that uses a marginal-adjusted null model: chance agreement is computed under the assumption that each expert rates "relevant" with probability equal to the observed marginal proportion. This is methodologically distinct from the modified kappa of Polit, Beck, and Owen (2007), which uses a fixed null (each expert independently rates relevant with probability 0.5). The two indices can therefore yield substantively different answers for the same data, particularly when the prevalence of "relevant" ratings is far from 0.5 (the typical case in content-validity work). Reporting both – alongside I-CVI – gives a more complete picture of inter-rater agreement than any single index. Wongpakaran et al. (2013, *BMC Medical Research Methodology*) recommended AC1 over Cohen's traditional kappa for high-prevalence rating contexts.

Usage

```
gwet_ac1(
  ratings,
  relevant_threshold = 3,
  na.rm = FALSE,
  ci = FALSE,
  n_boot = 2000,
  ci_method = c("percentile", "bca"),
  conf_level = 0.95,
  seed = NULL
)
```

Arguments

`ratings`	A numeric matrix or data frame of expert ratings (rows = experts, columns = items). A numeric vector is also accepted, treated as a single item.
`relevant_threshold`	Integer. Minimum rating considered "relevant". Ratings are dichotomized at this threshold before AC1 is computed, following standard practice in content-validity work (Polit, Beck, & Owen, 2007). Defaults to 3.
`na.rm`	Logical. If `TRUE`, missing ratings are excluded when counting experts. Defaults to `FALSE`.

ci	Logical. If TRUE, returns a data frame with bootstrap confidence intervals alongside the point estimate. Defaults to FALSE.
n_boot	Integer. Number of bootstrap replicates when ci = TRUE. Defaults to 2000 (Davison & Hinkley, 1997; Hesterberg, 2015).
ci_method	Character. One of "percentile" (default; Efron & Tibshirani, 1993) or "bca" (bias-corrected and accelerated; DiCiccio & Efron, 1996).
conf_level	Numeric. Confidence level between 0 and 1. Defaults to 0.95.
seed	Integer or NULL. If supplied, passed to set.seed() for reproducible bootstrap samples. Defaults to NULL.

Details

Optional bootstrap confidence intervals are available via ci = TRUE. Resampling is performed at the expert (row) level, matching the standard inferential frame for inter-rater reliability analyses (Gwet, 2014).

The formula is:

$$\text{AC1} = (p_a - p_e)/(1 - p_e)$$

For a single item with N experts of whom n_R rate as relevant:

$$p_a = [n_R(n_R - 1) + (N - n_R)(N - n_R - 1)]/[N(N - 1)]$$
$$p_e = 2\pi(1 - \pi), \quad \pi = n_R/N$$

This is Gwet's binary-rating form (Gwet, 2008, equation 5). The chance agreement term $p_e = 2\pi(1 - \pi)$ is maximised at 0.5 when $\pi = 0.5$ and approaches zero as π approaches either extreme.

Note that the "kappa paradox" (Feinstein & Cicchetti, 1990) and the Wongpakaran et al. (2013) comparison both refer to *Cohen's* kappa, whose chance-agreement term $\pi^2 + (1 - \pi)^2$ approaches 1 at the prevalence extremes. The modified kappa of Polit et al. (2007), implemented in this package as mod_kappa(), uses a different chance-correction ($C(N, A) \times 0.5^N$, a fixed binomial null) and does not behave like Cohen's kappa under high prevalence. The practical consequence is that mod_kappa and AC1 typically diverge when prevalence is far from 0.5 – modified kappa approaches I-CVI while AC1 discounts more of the observed agreement as prevalence-driven. Both are defensible; they answer different questions about chance.

Common interpretation cutoffs follow Altman (1991), as adapted to AC1 by Wongpakaran et al. (2013):

- AC1 < 0.20: poor
- AC1 0.20-0.39: fair
- AC1 0.40-0.59: moderate
- AC1 0.60-0.80: good
- AC1 > 0.80: very good

(Boundary values fall in the higher tier, matching the classifier used by apa_table() with interpretation_index = "gwet_ac1".)

Value

When `ci = FALSE` (default), a named numeric vector of AC1 values, one per item (or a single numeric value if `ratings` is a vector). When `ci = TRUE`, a data frame with columns `item`, `gwet_ac1`, `ci_lower`, `ci_upper`, `ci_method`, `conf_level`, `n_boot`.

References

Altman, D. G. (1991). *Practical statistics for medical research*. Chapman and Hall.

Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology*, 43(6), 543-549. doi:10.1016/08954356(90)90158L

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(1), 29-48. doi:10.1348/000711006X126600

Gwet, K. L. (2014). *Handbook of inter-rater reliability* (4th ed.). Advanced Analytics, LLC.

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*, 30(4), 459-467. doi:10.1002/nur.20199

Wongpakaran, N., Wongpakaran, T., Wedding, D., & Gwet, K. L. (2013). A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: A study conducted with personality disorder samples. *BMC Medical Research Methodology*, 13(1), 61. doi:10.1186/1471-22881361

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge University Press. doi:10.1017/CBO9780511802843

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, 11(3), 189-228. doi:10.1214/ss/1032280214

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman and Hall. doi:10.1201/9780429246593

Hesterberg, T. C. (2015). What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum. *The American Statistician*, 69(4), 371-386. doi:10.1080/00031305.2015.1089789

See Also

`mod_kappa()`, `icvi()`

Examples

```
ratings <- matrix(
  c(4, 4, 3, 4, 4,    # 5 of 5 relevant
    3, 4, 4, 4, 3,    # 5 of 5 relevant
    2, 3, 3, 4, 3,    # 4 of 5 relevant
    1, 2, 3, 2, 3),   # 2 of 5 relevant
  nrow = 5,
  dimnames = list(NULL, paste0("item", 1:4))
)
gwet_ac1(ratings)

# Compare with modified kappa to see Gwet's advantage at extremes
```

```
mod_kappa(ratings)

# With bootstrap confidence intervals
gwet_ac1(ratings, ci = TRUE, n_boot = 1000, seed = 1)
```

---

gwet_ac2 *Gwet's AC2 - weighted chance-corrected agreement for ordinal ratings*

Description

Computes Gwet's AC2 coefficient (Gwet, 2008, 2014) for ordinal ratings, which generalizes AC1 (see gwet_ac1()) to the case where rating categories are ordered and partial agreement between adjacent categories should count. Where AC1 dichotomizes ratings before computing chance- corrected agreement, AC2 preserves the full ordinal information through a weight matrix that assigns higher weights to pairs of ratings that are close together (e.g., a rating of 3 and 4) and lower weights to pairs that are far apart (e.g., 1 and 4).

Usage

```
gwet_ac2(
  ratings,
  weights = c("quadratic", "linear", "identity"),
  categories = NULL,
  na.rm = FALSE,
  ci = FALSE,
  n_boot = 2000,
  ci_method = c("percentile", "bca"),
  conf_level = 0.95,
  seed = NULL
)
```

Arguments

ratings	A numeric matrix or data frame of expert ratings (rows = experts, columns = items). A numeric vector is also accepted, treated as a single item.
weights	One of "quadratic" (default), "linear", "identity", or a custom $q \times q$ numeric weight matrix. Quadratic weights emphasize closeness between rating categories more strongly than linear weights. Identity weights reduce AC2 to AC1 on the raw (non-dichotomized) categories.
categories	Numeric vector of all possible rating values. Strongly recommended for content-validity work, where some categories may not appear in a given dataset. If NULL (the default), categories are inferred from the observed ratings, which can silently produce incorrect AC2 values when extreme categories are unused. See Details.

na.rm	Logical. If TRUE, missing ratings are excluded when counting experts on a per-item basis. Defaults to FALSE.
ci	Logical. If TRUE, returns a data frame with bootstrap confidence intervals alongside the point estimate. Defaults to FALSE.
n_boot	Integer. Number of bootstrap replicates when ci = TRUE. Defaults to 2000 (Davison & Hinkley, 1997; Hesterberg, 2015).
ci_method	Character. One of "percentile" (default; Efron & Tibshirani, 1993) or "bca" (bias-corrected and accelerated; DiCiccio & Efron, 1996).
conf_level	Numeric. Confidence level between 0 and 1. Defaults to 0.95.
seed	Integer or NULL. If supplied, passed to set.seed() for reproducible bootstrap samples. Defaults to NULL.

Details

Optional bootstrap confidence intervals are available via ci = TRUE. Resampling is performed at the expert (row) level, matching the standard inferential frame for inter-rater reliability analyses (Gwet, 2014).

For a single item with N experts whose ratings populate the q-category counts n_k ($k = 1, \ldots, q$) and weight matrix $W = (w_{kl})$:

$$p_a = \sum_k n_k(n_k^W - 1)/[N(N-1)]$$

where $n_k^W = \sum_l w_{kl} n_l$ is the weighted count for category k. Chance agreement uses Gwet's marginal-adjusted null:

$$p_e = T_w \sum_k \pi_k(1 - \pi_k)$$

with $T_w = \sum_{k,l} w_{kl}/[q(q-1)]$ and $\pi_k = n_k/N$. The coefficient is $\mathrm{AC2} = (p_a - p_e)/(1 - p_e)$.

This implementation reproduces the formulas used by the irrCAC package (by Kilem Gwet, the original author of AC1/AC2) so that AC2 values from this function are bit-for-bit equivalent to those from gwet.ac1.raw() from irrCAC on the same data with the same weight matrix and category list.

Quadratic and linear weights are computed as in Gwet (2014):

$$w_{kl}^{quad} = 1 - (c_k - c_l)^2/(c_q - c_1)^2$$
$$w_{kl}^{lin} = 1 - |c_k - c_l|/|c_q - c_1|$$

where $c_1, \ldots, c_q$ are the (sorted) category values.

Important: the categories argument should typically be set explicitly to the full theoretical rating scale (e.g., categories = 1:4 for a standard relevance scale), not left at NULL. If a particular item's ratings happen to use only a subset of categories (e.g., all experts rated 3 or 4), the default category-inference logic will produce a smaller weight matrix and substantially different AC2 values. This caveat matches the documented behavior of gwet.ac1.raw() from the irrCAC package.

Value

When `ci = FALSE` (default), a named numeric vector of AC2 values, one per item (or a single numeric value if `ratings` is a vector). When `ci = TRUE`, a data frame with columns `item`, `gwet_ac2`, `ci_lower`, `ci_upper`, `ci_method`, `conf_level`, `n_boot`.

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(1), 29-48. doi:10.1348/000711006X126600

Gwet, K. L. (2014). *Handbook of inter-rater reliability* (4th ed.). Advanced Analytics, LLC.

Wongpakaran, N., Wongpakaran, T., Wedding, D., & Gwet, K. L. (2013). A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: A study conducted with personality disorder samples. *BMC Medical Research Methodology*, 13(1), 61. doi:10.1186/1471-22881361

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge University Press. doi:10.1017/CBO9780511802843

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, 11(3), 189-228. doi:10.1214/ss/1032280214

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman and Hall. doi:10.1201/9780429246593

Hesterberg, T. C. (2015). What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum. *The American Statistician*, 69(4), 371-386. doi:10.1080/00031305.2015.1089789

See Also

`gwet_ac1()`, `mod_kappa()`

Examples

```
# Standard 4-point relevance scale, 5 experts on 4 items
ratings <- matrix(
  c(4, 4, 3, 4, 4,
    3, 4, 4, 4, 3,
    2, 3, 3, 4, 3,
    1, 2, 3, 2, 3),
  nrow = 5,
  dimnames = list(NULL, paste0("item", 1:4))
)

# Quadratic weights are the default and most common choice for
# ordinal data. Pass the full rating scale explicitly.
gwet_ac2(ratings, categories = 1:4)

# Linear weights are an alternative
gwet_ac2(ratings, weights = "linear", categories = 1:4)

# With bootstrap confidence intervals
```

```
gwet_ac2(ratings, categories = 1:4, ci = TRUE,
         n_boot = 1000, seed = 1)
```

---

icvi *Item-level Content Validity Index (I-CVI)*

Description

Computes the Item-level Content Validity Index (I-CVI) for one or more items rated by a panel of experts on a relevance scale. Following Lynn (1986) and Polit & Beck (2006), I-CVI is calculated as the proportion of experts who rate an item as 3 (relevant) or 4 (highly relevant) on a 4-point relevance scale.

Usage

```
icvi(
  ratings,
  relevant_threshold = 3,
  na.rm = FALSE,
  ci = FALSE,
  n_boot = 2000,
  ci_method = c("percentile", "bca"),
  conf_level = 0.95,
  seed = NULL
)
```

Arguments

ratings	A numeric matrix or data frame of expert ratings, where rows represent experts and columns represent items. Values are typically on a 1-4 relevance scale. A numeric vector is also accepted, treated as a single item.
relevant_threshold	Integer. The minimum rating considered "relevant". Defaults to 3 (i.e., ratings of 3 or 4 count as relevant on a 4-point scale).
na.rm	Logical. If TRUE, missing ratings are excluded from the calculation. Defaults to FALSE, in which case any NA produces NA for the affected item.
ci	Logical. If TRUE, returns a data frame with bootstrap confidence intervals in addition to the point estimate. Defaults to FALSE (returns a numeric vector, identical to the package's pre-0.2.0 behaviour).
n_boot	Integer. Number of bootstrap replicates when ci = TRUE. Defaults to 2000, following Davison and Hinkley (1997, ch. 5), who recommend at least 1000 replicates for stable percentile intervals, and Hesterberg (2015), who notes that 1000 is sufficient and 10,000 is ideal on modern hardware. 2000 balances stability against compute time.

ci_method	Character. One of "percentile" (default) or "bca" (bias-corrected and accelerated). Percentile (Efron & Tibshirani, 1993) respects the [0, 1] bounds of I-CVI naturally. BCa (DiCiccio & Efron, 1996) is preferred when the bootstrap distribution is skewed, which is common for I-CVI values near 1.0.
conf_level	Numeric. Confidence level between 0 and 1. Defaults to 0.95.
seed	Integer or NULL. If supplied, passed to set.seed() for reproducible bootstrap samples. Defaults to NULL.

Details

Optional bootstrap confidence intervals are available via ci = TRUE. When requested, the function resamples experts (rows) with replacement and recomputes I-CVI on each replicate. Resampling experts (rather than items) matches the standard inferential frame for inter-rater reliability analyses: experts are the random sample from a population of potential raters, while items are fixed by the study design (Gwet, 2014).

Common interpretation guidelines (Polit & Beck, 2006):

- I-CVI >= 0.78: excellent content validity (with 6 or more experts).
- I-CVI 0.70-0.78: acceptable, item may need revision.
- I-CVI < 0.70: item should be revised or eliminated.

With fewer than six experts, Lynn (1986) recommends a stricter cutoff of I-CVI = 1.00 for unanimous agreement.

Value

When ci = FALSE (default), a named numeric vector of I-CVI values, one per item (or a single numeric value if ratings is a vector). When ci = TRUE, a data frame with one row per item and columns item, icvi, ci_lower, ci_upper, ci_method, conf_level, n_boot.

References

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382-385. doi:10.1097/00006199198611000000017

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29(5), 489-497. doi:10.1002/nur.20147

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge University Press. doi:10.1017/CBO9780511802843

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, 11(3), 189-228. doi:10.1214/ss/1032280214

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman and Hall. doi:10.1201/9780429246593

Gwet, K. L. (2014). *Handbook of inter-rater reliability* (4th ed.). Advanced Analytics, LLC.

Hesterberg, T. C. (2015). What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum. *The American Statistician*, 69(4), 371-386. doi:10.1080/00031305.2015.1089789

Examples

```
# Five experts rating four items on a 1-4 relevance scale
ratings <- matrix(
  c(4, 4, 3, 4, 4,    # Item 1
    3, 4, 4, 4, 3,    # Item 2
    2, 3, 3, 4, 3,    # Item 3
    1, 2, 3, 2, 3),   # Item 4
  nrow = 5,
  dimnames = list(NULL, paste0("item", 1:4))
)
icvi(ratings)

# Single item supplied as a vector
icvi(c(4, 4, 3, 3, 4))

# Stricter threshold (only highest rating counts as relevant)
icvi(ratings, relevant_threshold = 4)

# With bootstrap confidence intervals (new in v0.2.0)
set.seed(1)
icvi(ratings, ci = TRUE, n_boot = 1000)

# BCa intervals, recommended when I-CVI values cluster near 1.0
icvi(ratings, ci = TRUE, ci_method = "bca", n_boot = 1000, seed = 1)
```

---

mod_kappa *Modified kappa - I-CVI adjusted for chance agreement*

Description

Computes modified kappa for each item, as proposed by Polit, Beck, and Owen (2007). Modified kappa adjusts the Item-level Content Validity Index (I-CVI) for chance agreement under the assumption that each expert independently rates an item as relevant with probability 0.5.

Usage

```
mod_kappa(
  ratings,
  relevant_threshold = 3,
  na.rm = FALSE,
  ci = FALSE,
  n_boot = 2000,
  ci_method = c("percentile", "bca"),
  conf_level = 0.95,
  seed = NULL
)
```

Arguments

ratings	A numeric matrix or data frame of expert ratings (rows = experts, columns = items). A numeric vector is also accepted, treated as a single item.
relevant_threshold	Integer. Minimum rating considered "relevant". Defaults to 3.
na.rm	Logical. If TRUE, missing ratings are excluded when counting experts and agreements. Defaults to FALSE.
ci	Logical. If TRUE, returns a data frame with bootstrap confidence intervals alongside the point estimate. Defaults to FALSE (returns a numeric vector, identical to the package's pre-0.2.0 behaviour).
n_boot	Integer. Number of bootstrap replicates when ci = TRUE. Defaults to 2000 (Davison & Hinkley, 1997; Hesterberg, 2015).
ci_method	Character. One of "percentile" (default; Efron & Tibshirani, 1993) or "bca" (bias-corrected and accelerated; DiCiccio & Efron, 1996).
conf_level	Numeric. Confidence level between 0 and 1. Defaults to 0.95.
seed	Integer or NULL. If supplied, passed to set.seed() for reproducible bootstrap samples. Defaults to NULL.

Details

Optional bootstrap confidence intervals are available via ci = TRUE. Resampling is performed at the expert (row) level, matching the standard inferential frame for inter-rater reliability analyses (Gwet, 2014).

The formula is:

$$\kappa^* = (\text{I-CVI} - P_c)/(1 - P_c)$$

where the chance agreement probability is

$$P_c = \binom{N}{A} \times 0.5^N$$

with N = number of experts and A = number of experts rating the item as relevant.

Common interpretation cutoffs (Cicchetti and Sparrow, 1981; adopted by Polit et al., 2007):

- kappa* < 0.40: poor
- kappa* 0.40-0.59: fair
- kappa* 0.60-0.74: good
- kappa* > 0.74: excellent

Value

When ci = FALSE (default), a named numeric vector of modified-kappa values, one per item (or a single numeric value if ratings is a vector). When ci = TRUE, a data frame with one row per item and columns item, mod_kappa, ci_lower, ci_upper, ci_method, conf_level, n_boot.

References

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, 86(2), 127-137.

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*, 30(4), 459-467. doi:10.1002/nur.20199

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge University Press. doi:10.1017/CBO9780511802843

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, 11(3), 189-228. doi:10.1214/ss/1032280214

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman and Hall. doi:10.1201/9780429246593

Gwet, K. L. (2014). *Handbook of inter-rater reliability* (4th ed.). Advanced Analytics, LLC.

Hesterberg, T. C. (2015). What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum. *The American Statistician*, 69(4), 371-386. doi:10.1080/00031305.2015.1089789

See Also

`icvi()`

Examples

```
ratings <- matrix(
  c(4, 4, 3, 4, 4,
    3, 4, 4, 4, 3,
    2, 3, 3, 4, 3,
    1, 2, 3, 2, 3),
  nrow = 5,
  dimnames = list(NULL, paste0("item", 1:4))
)
mod_kappa(ratings)

# With bootstrap confidence intervals (new in v0.2.0)
mod_kappa(ratings, ci = TRUE, n_boot = 1000, seed = 1)
```

---

plot.content_validity *Plot a content validity analysis*

Description

Produces an I-CVI / chance-corrected agreement scatter plot for the item-level results of a `content_validity()` analysis, parallel to the difficulty-discrimination scatter used in classical item analysis. Items that fall outside the conventional adequacy region are flagged in red and labeled by default.

Usage

```
## S3 method for class 'content_validity'
plot(
  x,
  y = NULL,
  y_index = c("mod_kappa", "gwet_ac1", "gwet_ac2", "aiken_v"),
  label = c("flagged", "all", "none"),
  flag_logic = c("any", "icvi", "y_index", "both"),
  flag_threshold_icvi = 0.78,
  flag_threshold_y = NULL,
  point_cex = 1.4,
  label_cex = 0.75,
  ...
)
```

Arguments

x	A content_validity object returned by content_validity().
y	Ignored (required by the S3 plot generic).
y_index	Character. Which agreement index to display on the y-axis. One of "mod_kappa" (default), "gwet_ac1", "gwet_ac2", or "aiken_v".
label	Character. One of "flagged" (default, label only items outside the adequacy region), "all", or "none".
flag_logic	Character. Which axis (or axes) drive the flagging. One of "any" (default; flag items below either threshold, useful for "items that need any review"), "icvi" (flag only items below the I-CVI threshold), "y_index" (flag only items below the y-axis threshold, useful when the plot is presenting one index specifically), or "both" (strict; flag only items below both thresholds).
flag_threshold_icvi	Numeric. Lower I-CVI threshold marking the adequacy region (Polit & Beck, 2006). Defaults to 0.78.
flag_threshold_y	Numeric. Lower threshold on the y-axis index. Defaults depend on y_index: 0.74 for mod_kappa (Cicchetti & Sparrow, 1981), 0.60 for AC1 and AC2 (Altman, 1991), 0.70 for Aiken's V (Aiken, 1985).
point_cex	Numeric. Point expansion factor. Default 1.4.
label_cex	Numeric. Label expansion factor. Default 0.75.
...	Currently ignored.

Value

Invisibly returns x. Called for its side effect (a base R plot drawn on the current graphics device).

References

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131-142. doi:10.1177/0013164485451012

Altman, D. G. (1991). *Practical statistics for medical research*. Chapman and Hall.

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items. *American Journal of Mental Deficiency*, 86(2), 127-137.

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? *Research in Nursing & Health*, 29(5), 489-497. doi:10.1002/nur.20147

Examples

```
data(cvi_example)
result <- content_validity(cvi_example)
plot(result)
plot(result, y_index = "gwet_ac2")
plot(result, y_index = "aiken_v", label = "all")
```

`print.content_validity` *Print method for content_validity objects*

Description

Print method for content_validity objects

Usage

```
## S3 method for class 'content_validity'
print(x, digits = 4, ...)
```

Arguments

`x`	A `content_validity` object returned by `content_validity()`.
`digits`	Integer. Number of digits to round numeric output to.
`...`	Currently ignored.

Value

Invisibly returns `x`.

scvi_ave *Scale-level Content Validity Index, Average method (S-CVI/Ave)*

Description

Computes the Scale-level Content Validity Index using the averaging method, defined as the mean of the Item-level Content Validity Indices (I-CVI) across all items in the instrument.

Usage

```
scvi_ave(
  ratings,
  relevant_threshold = 3,
  na.rm = FALSE,
  ci = FALSE,
  n_boot = 2000,
  ci_method = c("percentile", "bca"),
  conf_level = 0.95,
  seed = NULL
)
```

Arguments

ratings	A numeric matrix or data frame of expert ratings (rows = experts, columns = items) on a relevance scale.
relevant_threshold	Integer. Minimum rating considered "relevant". Defaults to 3.
na.rm	Logical. Passed through to icvi(). Defaults to FALSE.
ci	Logical. If TRUE, returns a data frame with a bootstrap confidence interval alongside the point estimate. Defaults to FALSE (returns a single numeric value, identical to the package's pre-0.2.0 behaviour).
n_boot	Integer. Number of bootstrap replicates when ci = TRUE. Defaults to 2000 (Davison & Hinkley, 1997; Hesterberg, 2015).
ci_method	Character. One of "percentile" (default; Efron & Tibshirani, 1993) or "bca" (bias-corrected and accelerated; DiCiccio & Efron, 1996).
conf_level	Numeric. Confidence level between 0 and 1. Defaults to 0.95.
seed	Integer or NULL. If supplied, passed to set.seed() for reproducible bootstrap samples. Defaults to NULL.

Details

Optional bootstrap confidence intervals are available via ci = TRUE. Resampling is performed at the expert (row) level, matching the standard inferential frame for inter-rater reliability analyses (Gwet, 2014).

S-CVI/Ave >= 0.90 is generally considered excellent content validity at the scale level (Polit & Beck, 2006). Note that S-CVI is undefined for a single item; supply a matrix or data frame with two or more item columns.

Value

When `ci = FALSE` (default), a single numeric value: the average I-CVI across items. When `ci = TRUE`, a one-row data frame with columns `item` (set to `"scale"`), `scvi_ave`, `ci_lower`, `ci_upper`, `ci_method`, `conf_level`, `n_boot`.

References

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29(5), 489-497. doi:10.1002/nur.20147

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge University Press. doi:10.1017/CBO9780511802843

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, 11(3), 189-228. doi:10.1214/ss/1032280214

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman and Hall. doi:10.1201/9780429246593

Gwet, K. L. (2014). *Handbook of inter-rater reliability* (4th ed.). Advanced Analytics, LLC.

Hesterberg, T. C. (2015). What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum. *The American Statistician*, 69(4), 371-386. doi:10.1080/00031305.2015.1089789

See Also

`icvi()`

Examples

```
ratings <- matrix(
  c(4, 4, 3, 4, 4,
    3, 4, 4, 4, 3,
    2, 3, 3, 4, 3,
    1, 2, 3, 2, 3),
  nrow = 5
)
scvi_ave(ratings)

# With bootstrap confidence interval (new in v0.2.0)
scvi_ave(ratings, ci = TRUE, n_boot = 1000, seed = 1)
```

---

scvi_ua *Scale-level Content Validity Index, Universal Agreement method (S-CVI/UA)*

Description

Computes the Scale-level Content Validity Index using the universal agreement method, defined as the proportion of items where all experts rate the item as relevant.

Usage

```
scvi_ua(
  ratings,
  relevant_threshold = 3,
  na.rm = FALSE,
  ci = FALSE,
  n_boot = 2000,
  ci_method = c("percentile", "bca"),
  conf_level = 0.95,
  seed = NULL
)
```

Arguments

ratings	A numeric matrix or data frame of expert ratings (rows = experts, columns = items) on a relevance scale.
relevant_threshold	Integer. Minimum rating considered "relevant". Defaults to 3.
na.rm	Logical. If TRUE, missing ratings are ignored when checking universal agreement. Defaults to FALSE.
ci	Logical. If TRUE, returns a data frame with a bootstrap confidence interval alongside the point estimate. Defaults to FALSE (returns a single numeric value, identical to the package's pre-0.2.0 behaviour).
n_boot	Integer. Number of bootstrap replicates when ci = TRUE. Defaults to 2000 (Davison & Hinkley, 1997; Hesterberg, 2015).
ci_method	Character. One of "percentile" (default; Efron & Tibshirani, 1993) or "bca" (bias-corrected and accelerated; DiCiccio & Efron, 1996).
conf_level	Numeric. Confidence level between 0 and 1. Defaults to 0.95.
seed	Integer or NULL. If supplied, passed to set.seed() for reproducible bootstrap samples. Defaults to NULL.

Details

Optional bootstrap confidence intervals are available via ci = TRUE. Resampling is performed at the expert (row) level, matching the standard inferential frame for inter-rater reliability analyses (Gwet, 2014).

S-CVI/UA is a stricter criterion than S-CVI/Ave and tends to produce lower values, especially with larger expert panels. Polit and Beck (2006) recommend reporting both indices together. With small panels of 3-5 experts, S-CVI/UA >= 0.80 is often considered acceptable.

Value

When ci = FALSE (default), a single numeric value: the proportion of items with universal agreement. When ci = TRUE, a one-row data frame with columns item (set to "scale"), scvi_ua, ci_lower, ci_upper, ci_method, conf_level, n_boot.

References

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29(5), 489-497. doi:10.1002/nur.20147

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge University Press. doi:10.1017/CBO9780511802843

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, 11(3), 189-228. doi:10.1214/ss/1032280214

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman and Hall. doi:10.1201/9780429246593

Gwet, K. L. (2014). *Handbook of inter-rater reliability* (4th ed.). Advanced Analytics, LLC.

Hesterberg, T. C. (2015). What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum. *The American Statistician*, 69(4), 371-386. doi:10.1080/00031305.2015.1089789

See Also

`icvi()`, `scvi_ave()`

Examples

```
ratings <- matrix(
  c(4, 4, 3, 4, 4,
    3, 4, 4, 4, 3,
    2, 3, 3, 4, 3,
    1, 2, 3, 2, 3),
  nrow = 5
)
scvi_ua(ratings)

# With bootstrap confidence interval (new in v0.2.0)
scvi_ua(ratings, ci = TRUE, n_boot = 1000, seed = 1)
```

Index

∗ **datasets**
cvi_example, 10

aiken_v, 2
aiken_v(), *6, 7*
apa_table, 4
apa_table(), *15*

content_validity, 6
content_validity(), *5*, *24–26*
cv_sample_size_icvi, 8
cvi_example, 10
cvr, 11
cvr(), *7*, *13*
cvr_critical, 13
cvr_critical(), *7*, *12*

gwet_ac1, 14
gwet_ac1(), *6, 7*, *17*, *19*
gwet_ac2, 17
gwet_ac2(), *6, 7*

icvi, 20
icvi(), *4*, *6, 7*, *9*, *16*, *24*, *27, 28*, *30*

knitr::kable(), *4*

mod_kappa, 22
mod_kappa(), *6, 7*, *15, 16*, *19*

plot.content_validity, 24
print.content_validity, 26

scvi_ave, 27
scvi_ave(), *6, 7*, *30*
scvi_ua, 28
scvi_ua(), *6, 7*
set.seed(), *3*, *11*, *15*, *18*, *21*, *23*, *27*, *29*
stats::uniroot(), *8, 9*