Pearson's r is a special case of Cohen's weighted kappa

Matthijs J Warrens
Leiden University

IFCS 2013

Classifying objects

Objects	Rater
1	A
2	B
3	A
4	C
5	C
6	B
$\vdots$	$\vdots$

- Objects are classified into 3 categories
- Categories ordinal (A, B, C)
- Categories defined in advance

Classifying objects

Objects	Rater
1	A
2	B
3	A
4	C
5	C
6	B
⋮	⋮

- Objects are classified into 3 categories
- Categories ordinal (A, B, C)
- Categories defined in advance
- Quality ratings?
  No golden standard
- Quality coding system?

Two raters

Objects	Rater 1	Rater 2
1	A	A
2	B	A
3	A	B
4	C	C
5	C	A
6	B	B
⋮	⋮	⋮

- Objects are classified into 3 categories by two raters

Cross-classification

Objects	Rater 1	Rater 2
1	A	A
2	B	A
3	A	B
4	C	C
5	C	A
6	B	B
⋮	⋮	⋮

	Rater 2			
Rater 1	A	B	C	Total
A	9	1	0	10
B	4	20	5	29
C	1	4	36	41
Total	14	25	41	80

Data example

- Glasgow outcome scale
  Anderson et al (1993)

- A = Severe disability
  B = Moderate disability
  C = Good recovery

	Rater 2			
Rater 1	A	B	C	Total
A	9	1	0	10
B	4	20	5	29
C	1	4	36	41
Total	14	25	41	80

Summarizing agreement

- Analyse agreement
  Log-linear models
- Summarize by a single number
  Widely used:
  weighted kappa
  (Cohen 1968)

	Rater 2			
Rater 1	A	B	C	Total
A	9	1	0	10
B	4	20	5	29
C	1	4	36	41
Total	14	25	41	80

Contingency table

- Elements: n_{jk}
- Marginal totals: $n_{j\cdot}$, $n_{\cdot j}$

	Rater 2			
Rater 1	A	B	C	Total
A	n_{11}	n_{12}	n_{13}	$n_{1\cdot}$
B	n_{21}	n_{22}	n_{23}	$n_{2\cdot}$
C	n_{31}	n_{32}	n_{33}	$n_{3\cdot}$
Total	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot 3}$	n

Definition

Cohen (1968), weights $w_{jk} \geq 0$

$$\kappa_w := 1 - \frac{n \sum_{j=1}^{m} \sum_{k=1}^{m} n_{jk} w_{jk}}{\sum_{j=1}^{m} \sum_{k=1}^{m} n_{j\cdot} n_{\cdot k} w_{jk}}$$

- $w_{jk} = \mathbb{1}_{j \neq k}$ (identity weights; Cohen, 1960)
- $w_{jk} = |j - k|$ (linear weights)
- $w_{jk} = (j - k)^2$ (quadratic weights)
- Arbitrary definitions

Quadratic weights

Quadratically weighted kappa

$$\kappa_q := 1 - \frac{n \sum_{j=1}^{m} \sum_{k=1}^{m} n_{jk}(j-k)^2}{\sum_{j=1}^{m} \sum_{k=1}^{m} n_{j\cdot} n_{\cdot k}(j-k)^2}$$

- Most widely used
- Fleiss and Cohen (1973): proportion of variance
- Schuster (2004): interpretation in terms of means and variances

Some more notation

	Rater 1	Rater 2
Objects	X	Y
1	x_1	y_1
2	x_2	y_2
3	x_3	y_3
4	x_4	y_4
5	x_5	y_5
6	x_6	y_6
$\vdots$	$\vdots$	$\vdots$

- Means: $\bar{x}$, $\bar{y}$
  Variances: s_x, s_y
  Covariance: s_{xy}
- $x_i, y_i \in \{z_1, \ldots, z_m\}$
  z_j is used for coding category j

Theorem

Define

$$\kappa_r := 1 - \frac{n \sum_{j=1}^{m} \sum_{k=1}^{m} n_{jk} \left(\frac{z_j - \bar{x}}{s_x} - \frac{z_k - \bar{y}}{s_y} \right)^2}{\sum_{j=1}^{m} \sum_{k=1}^{m} n_{j.} n_{.k} \left(\frac{z_j - \bar{x}}{s_x} - \frac{z_k - \bar{y}}{s_y} \right)^2}$$

and

$$r := \frac{s_{xy}}{s_x s_y}.$$

Then

$$\kappa_r = r.$$

Identity for numerator

$$\sum_{j=1}^{m}\sum_{k=1}^{m} n_{jk}\left(\frac{z_j-\bar{x}}{s_x}-\frac{z_k-\bar{y}}{s_y}\right)^2=\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s_x}-\frac{y_i-\bar{y}}{s_y}\right)^2$$

$$=\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s_x}\right)^2+\sum_{i=1}^{n}\left(\frac{y_i-\bar{y}}{s_y}\right)^2-2\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s_x}\right)\left(\frac{y_i-\bar{y}}{s_y}\right)$$

$$=\frac{(n-1)s_x^2}{s_x^2}+\frac{(n-1)s_y^2}{s_y^2}-\frac{2(n-1)s_{xy}}{s_x s_y}$$

$$=2(n-1)-2(n-1)r=2(n-1)(1-r)$$

Identity for denominator

$$\frac{1}{n}\sum_{j=1}^{m}\sum_{k=1}^{m} n_{j.}n_{.k}\left(\frac{z_j-\bar{x}}{s_x}-\frac{z_k-\bar{y}}{s_y}\right)^2$$

$$=\frac{1}{n}\sum_{j=1}^{m} n_{j.}\left(\frac{z_j-\bar{x}}{s_x}\right)^2\sum_{k=1}^{m} n_{.k}+\frac{1}{n}\sum_{j=1}^{m} n_{j.}\sum_{k=1}^{m} n_{.k}\left(\frac{z_k-\bar{y}}{s_y}\right)^2-\frac{2}{n}\sum_{j=1}^{m} n_{j.}\left(\frac{z_j-\bar{x}}{s_x}\right)\sum_{k=1}^{m} n_{.k}\left(\frac{z_k-\bar{y}}{s_y}\right)$$

$$=\sum_{j=1}^{m} n_{j.}\left(\frac{z_j-\bar{x}}{s_x}\right)^2+\sum_{k=1}^{m} n_{.k}\left(\frac{z_k-\bar{y}}{s_y}\right)^2=\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s_x}\right)^2+\sum_{i=1}^{n}\left(\frac{y_i-\bar{y}}{s_y}\right)^2=2(n-1)$$

using identities

$$\sum_{j=1}^{m} n_{j.}=n \qquad \sum_{j=1}^{m} n_{j.}z_j=\sum_{i=1}^{n} x_i=n\bar{x} \qquad \sum_{j=1}^{m} n_{j.}z_j^2=\sum_{i=1}^{n} x_i^2$$

Hence

$$\kappa_r=1-\frac{2(n-1)(1-r)}{2(n-1)}=r.$$

Conclusion

- Weighted kappa widely used
- Weights are in general arbitrarily defined
- If weights may depend on the data, Pearson's r a special case, which has a clear interpretation
- Weighted kappa coefficients obsolete?

References

- Anderson SI et al (1993) Glasgow outcome scale: An inter-rater reliability study. Brain Injury 7:309-317.
- 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 JL, Cohen J (1973) The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educational and Psychological Measurement 33:613-619.
- Schuster C (2004) A note on the interpretation of weighted kappa and its relations to other rater agreement statistics for metric scales. Educational and Psychological Measurement 64:243-253.
- Warrens MJ (2014) Corrected Zegers-ten Berge coefficients are special cases of Cohen's weighted kappa. Journal of Classification.