



Universiteit  
Leiden  
The Netherlands

## Similarity coefficients for binary data : properties of coefficients, coefficient matrices, multi-way metrics and multivariate coefficients

Warrens, M.J.

### Citation

Warrens, M. J. (2008, June 25). *Similarity coefficients for binary data : properties of coefficients, coefficient matrices, multi-way metrics and multivariate coefficients*. Retrieved from <https://hdl.handle.net/1887/12987>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/12987>

**Note:** To cite this publication please use the final published version (if applicable).

## STELLINGEN

Behorend bij het proefschrift *Similarity Coefficients for Binary Data*

1. Men kan de Hubert-Arabie adjusted Rand index uitrekenen door eerst de  $2 \times 2$  tabel te formeren door het aantal objectparen te tellen die in hetzelfde cluster zijn geplaatst door beide methodes, die in een cluster zijn geplaatst door een methode maar in verschillende clusters door de andere methode, en in verschillende clusters door beide methodes, en vervolgens Cohen's kappa uit te rekenen (dit proefschrift).
2. De simple matching coëfficiënt, Cohen's kappa, Goodman en Kruskal's lambda, Scott's pi, Hamann's eta, en overeenstemmingsmaten geïntroduceerd door Dice/ Sørensen en Rogot en Goldberg, worden equivalent na correctie voor kans, ongeacht de verwachte waarde die gebruikt wordt (dit proefschrift).
3. De Loevinger coëfficiënt is de enige lineaire transformatie van de geobserveerde proportie van overeenstemming met een waarde 0 als de twee variabelen statistisch onafhankelijk zijn en een maximale waarde van 1 ongeacht de verdeling van de marginalen (dit proefschrift).
4. Gebruik Yule's  $Q$  of Yule's  $Y$ , beide niet-lineaire transformaties van de odds ratio, als het van belang is dat een coëfficiënt een waarde 0 heeft als de twee variabelen statistisch onafhankelijk zijn, en een maximale waarde van 1 en een minimale waarde van -1 heeft, ongeacht de verdeling van de marginalen. Er bestaat geen lineaire transformatie van de geobserveerde proportie van overeenstemming die aan deze drie condities voldoet (dit proefschrift).
5. Hoewel een nieuw resultaat in een onderzoeksveld waar al vele onderzoekers aandacht en tijd aan hebben besteed, meer indruk maakt, is voor een nieuw resultaat in een onderzoeksveld waar weinig tot geen onderzoekers aandacht aan hebben besteed, meer fantasie nodig.
6. Het gezegde "een plaatje zegt meer dan duizend woorden" geeft precies de subjectiviteit bij het interpreteren van plaatjes weer. In de exploratieve data-analyse zou dit reden genoeg moeten zijn om voor eens en altijd te stoppen met het evalueren van visualisatietechnieken en modellen aan de hand van plaatjes.
7. Juist omdat het zo mooi is om dingen te durven denken die nog nooit gedacht zijn, is het zo betreurenswaardig dat bij zoveel sociaalwetenschappelijk promotieonderzoek de inhoud van het proefschrift al grotendeels vastligt alvorens de promovendus begonnen is.
8. In plaats van vijf tijdschriftartikelen te bundelen met een nietje, is het in ieder geval veel leuker, maar misschien ook wel beter voor je ontwikkeling als beginnend wetenschapper, om vier jaar lang alles wat in je hoofd opkomt, wel of niet gerelateerd aan het onderzoeksvoorstel, te exploreren en op te schrijven.
9. Mensen vallen grofweg te verdelen in pratere en luisteraars, waarbij de eerste groep weer onderverdeeld kan worden in mensen die gewoon hard praten en mensen die iets te zeggen hebben.
10. Als je gedurende een relatie nooit cadeautjes geeft of bloemen meebrengt, krijg je in een later stadium ook niet het verwijt dat je vroeger veel attenter was.

Leiden, mei 2008  
Matthijs J. Warrens