



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).

# Similarity Coefficients for Binary Data

Warrens, Matthijs Joost,  
Similarity Coefficients for Binary Data. Properties of Coefficients, Coefficient  
Matrices, Multi-way Metrics and Multivariate Coefficients  
Dissertation Leiden University – With References – With Summary in Dutch.

Subject headings: Association Measures; Correction for Chance; Correction for  
Maximum Value; Homogeneity Analysis;  $k$ -Way Metricity.

ISBN 978-90-8891-0524

© 2008 Matthijs J. Warrens

Printed by Proefschriftmaken.nl, Oisterwijk

# Similarity Coefficients for Binary Data

*Properties of Coefficients, Coefficient Matrices,  
Multi-way Metrics and Multivariate Coefficients*

PROEFSCHRIFT

ter verkrijging van de graad van Doctor aan de Universiteit Leiden,  
op gezag van Rector Magnificus prof. mr. P.F. van der Heijden,  
volgens besluit van het College voor Promoties  
te verdedigen op woensdag 25 juni 2008  
klokke 13.45 uur

door

Matthijs Joost Warrens

geboren te Rotterdam  
in 1978

## PROMOTIECOMMISSIE

Promotor	Prof. dr. W.J. Heiser
Co-promotor	Dr. D.N.M. de Gruijter
Referent	Prof. dr. J.C. Gower, The Open University, Milton Keynes, UK
Overige Leden	Prof. dr. J.J. Meulman Prof. dr. H.A.L. Kiers, University of Groningen Dr. M. de Rooij

*To Sascha*

## Acknowledgments

There are several people I would like to thank for their help in one way or another during the time of writing this dissertation: Laurence Frank, for showing me around when I took my first steps as a PhD student, and for the many conversations on the various aspects of research; and Marike Polak, with whom I shared a room for such a substantial period, for her companionship and the many talks on Dutch soccer and life in general. I also thank Marian Hickendorff and Susa~ña Verdel for just being there.

---

# Contents

---

<b>Prologue</b>	<b>xi</b>
<b>I Similarity coefficients</b>	<b>1</b>
<b>1 Coefficients for binary variables</b>	<b>3</b>
1.1 Four dependent quantities . . . . .	4
1.2 Axioms for (dis)similarities . . . . .	7
1.3 Uncorrelatedness and statistical independence . . . . .	10
1.4 Indeterminacy . . . . .	12
1.5 Epilogue . . . . .	17
<b>2 Coefficients for nominal and quantitative variables</b>	<b>19</b>
2.1 Nominal variables . . . . .	20
2.2 Comparing two partitions . . . . .	22
2.3 Comparing two judges . . . . .	24
2.4 Quantitative variables . . . . .	25
2.5 Measures from set theory . . . . .	27
2.6 Epilogue . . . . .	28
<b>3 Coefficient families</b>	<b>29</b>
3.1 Parameter families . . . . .	30
3.2 Power means . . . . .	35
3.3 A general family . . . . .	37
3.4 Linearity . . . . .	39
3.5 Epilogue . . . . .	41



<b>4</b>	<b>Correction for chance agreement</b>	<b>43</b>
4.1	Some equivalences . . . . .	44
4.2	Expectations . . . . .	48
4.3	Two transformations . . . . .	51
4.4	Corrected coefficients . . . . .	52
4.5	Epilogue . . . . .	55
<b>5</b>	<b>Correction for maximum value</b>	<b>57</b>
5.1	Maximum value . . . . .	58
5.2	Correction for maximum value . . . . .	60
5.3	Correction for minimum value . . . . .	62
5.4	Epilogue . . . . .	65
5.5	Loevinger's coefficient . . . . .	66
<b>II</b>	<b>Similarity matrices</b>	<b>69</b>
<b>6</b>	<b>Data structures</b>	<b>71</b>
6.1	Latent variable models . . . . .	72
6.2	Petrie structure . . . . .	74
6.3	Guttman items . . . . .	77
6.4	Epilogue . . . . .	80
<b>7</b>	<b>Robinson matrices</b>	<b>81</b>
7.1	Auxiliary results . . . . .	82
7.2	Braun-Blanquet + Russel and Rao coefficient . . . . .	83
7.3	Double Petrie . . . . .	84
7.4	Restricted double Petrie . . . . .	85
7.5	Counterexamples . . . . .	86
7.6	Epilogue . . . . .	87
<b>8</b>	<b>Eigenvector properties</b>	<b>89</b>
8.1	Ordered eigenvector elements . . . . .	90
8.2	Related eigenvectors . . . . .	93
8.3	Homogeneity analysis . . . . .	94
8.4	Epilogue . . . . .	98
<b>9</b>	<b>Homogeneity analysis and the 2-parameter IRT model</b>	<b>99</b>
9.1	Classical item analysis . . . . .	100
9.2	Person parameter . . . . .	101
9.3	Discrimination parameter . . . . .	102
9.4	More discrimination parameters . . . . .	105
9.5	Location parameter and category weights . . . . .	107
9.6	Epilogue . . . . .	108

<b>10 Metric properties of two-way coefficients</b>	<b>109</b>
10.1 Dissimilarity coefficients . . . . .	110
10.2 Main results . . . . .	111
10.3 Counterexamples . . . . .	114
10.4 Epilogue . . . . .	115
<b>III Multi-way metrics</b>	<b>117</b>
<b>11 Axiom systems</b>	<b>119</b>
11.1 Two-way dissimilarities . . . . .	120
11.2 Three-way dissimilarities . . . . .	122
11.3 Multi-way dissimilarities . . . . .	127
11.4 Epilogue . . . . .	130
<b>12 Multi-way metrics</b>	<b>131</b>
12.1 Definitions . . . . .	132
12.2 Two identical objects . . . . .	134
12.3 Bounds . . . . .	136
12.4 $(k - 1)$ -Way metrics implied by $k$ -way metrics . . . . .	139
12.5 Epilogue . . . . .	142
<b>13 Multi-way ultrametries</b>	<b>143</b>
13.1 Definitions . . . . .	144
13.2 Strong ultrametries . . . . .	145
13.3 More strong ultrametries . . . . .	148
13.4 Metrics implied by ultrametries . . . . .	149
13.5 Epilogue . . . . .	150
<b>14 Perimeter models</b>	<b>151</b>
14.1 Definitions . . . . .	152
14.2 Decompositions . . . . .	153
14.3 Metric properties . . . . .	155
14.4 Maximum distance . . . . .	156
14.5 Epilogue . . . . .	158
<b>15 Generalizations of Theorem 10.3</b>	<b>159</b>
15.1 A generalization of Theorem 10.3. . . . .	160
15.2 Auxiliary results . . . . .	162
15.3 A stronger generalization of Theorem 10.3 . . . . .	164
15.4 Epilogue . . . . .	167

<b>IV</b>	<b>Multivariate coefficients</b>	<b>169</b>
<b>16</b>	<b>Coefficients that generalize basic characteristics</b>	<b>171</b>
16.1	Bennani-Heiser coefficients . . . . .	172
16.2	Dice's association indices . . . . .	175
16.3	Bounds . . . . .	178
16.4	Epilogue . . . . .	179
<b>17</b>	<b>Multi-way coefficients based on two-way quantities</b>	<b>181</b>
17.1	Multivariate formulations . . . . .	182
17.2	Main results . . . . .	184
17.3	Gower-Legendre families . . . . .	185
17.4	Bounds . . . . .	187
17.5	Epilogue . . . . .	188
<b>18</b>	<b>Metric properties of multivariate coefficients</b>	<b>191</b>
18.1	Russel-Rao coefficient . . . . .	192
18.2	Simple matching coefficient . . . . .	194
18.3	Jaccard coefficient . . . . .	196
18.4	Epilogue . . . . .	198
<b>19</b>	<b>Robinson cubes</b>	<b>199</b>
19.1	Definitions . . . . .	200
19.2	Functions . . . . .	202
19.3	Coefficient properties . . . . .	204
19.4	Epilogue . . . . .	205
	<b>References</b>	<b>207</b>
	<b>List of similarity coefficients</b>	<b>219</b>
	<b>Summary of coefficient properties</b>	<b>223</b>
	<b>Coefficient index</b>	<b>227</b>
	<b>Author index</b>	<b>229</b>
	<b>Summary in Dutch (Samenvatting)</b>	<b>233</b>
	<b>Curriculum vitae</b>	<b>237</b>