

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

Prologue	xi
I Similarity coefficients	1
1 Coefficients for binary variables	3
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
2 Coefficients for nominal and quantitative variables	19
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
3 Coefficient families	29
3.1 Parameter families	30
3.2 Power means	35
3.3 A general family	37
3.4 Linearity	39
3.5 Epilogue	41

4	Correction for chance agreement	43
4.1	Some equivalences	44
4.2	Expectations	48
4.3	Two transformations	51
4.4	Corrected coefficients	52
4.5	Epilogue	55
5	Correction for maximum value	57
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
II	Similarity matrices	69
6	Data structures	71
6.1	Latent variable models	72
6.2	Petrie structure	74
6.3	Guttman items	77
6.4	Epilogue	80
7	Robinson matrices	81
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
8	Eigenvector properties	89
8.1	Ordered eigenvector elements	90
8.2	Related eigenvectors	93
8.3	Homogeneity analysis	94
8.4	Epilogue	98
9	Homogeneity analysis and the 2-parameter IRT model	99
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

10 Metric properties of two-way coefficients	109
10.1 Dissimilarity coefficients	110
10.2 Main results	111
10.3 Counterexamples	114
10.4 Epilogue	115
III Multi-way metrics	117
11 Axiom systems	119
11.1 Two-way dissimilarities	120
11.2 Three-way dissimilarities	122
11.3 Multi-way dissimilarities	127
11.4 Epilogue	130
12 Multi-way metrics	131
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
13 Multi-way ultrametries	143
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
14 Perimeter models	151
14.1 Definitions	152
14.2 Decompositions	153
14.3 Metric properties	155
14.4 Maximum distance	156
14.5 Epilogue	158
15 Generalizations of Theorem 10.3	159
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

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