
Contents

Acknowledgments	ix
Introduction	xi
I.1. Prerequisites and notations	xi
I.2. The case under study is a genomic monster	xvii
Chapter 1. Introduction to Correspondence	
Analysis	1
1.1. Chapter objectives	1
1.2. Metric choice	2
1.2.1. A small data set example	2
1.2.2. Euclidean distance	7
1.2.3. Euclidean distance on row profiles	8
1.2.4. Euclidean distance on double profiles	9
1.2.5. The χ^2 distance	10
1.3. Properties	12
1.3.1. Scree plot of eigenvalues	12
1.3.2. Factor orientations are meaningless	15
1.3.3. CA is symmetric	17
1.3.4. Nonlinearity in CA is not a bug but a feature	17
1.3.5. Distributional equivalence principle	20
1.3.6. So, why the χ^2 metric?	20

Chapter 2. Global Correspondence Analysis	25
2.1. Data set	25
2.1.1. Queries in database	25
2.1.2. Data polishing	31
2.2. Running global correspondence analysis	34
2.3. The missing factor F_0	38
2.4. First factor	40
2.4.1. Coding sequence point of view	40
2.4.2. Codon point of view	42
2.4.3. Biological interpretation	44
2.5. Second and third factors	44
2.5.1. Coding sequence point of view	45
2.5.2. Codon point of view	48
2.5.3. Biological interpretation	51
2.6. Fourth and fifth factors	53
2.6.1. Coding sequence point of view	55
2.6.2. Codon point of view	55
Chapter 3. Within and Between Correspondence Analysis	59
3.1. Running the analyses	59
3.2. Synonymous codon usage (WCA)	63
3.2.1. The first and unique factor F_1	63
3.2.1.1. Coding sequences point of view	63
3.2.1.2. Codon point of view	63
3.2.1.3. Biological interpretation	65
3.3. Amino acid usage (BCA)	66
3.3.1. The missing factor F_0	66
3.3.2. The ugly (F_1, F_2, F_3) <i>ménage à trois</i>	66
3.3.2.1. Coding sequences point of view	66
3.3.2.2. Aminoacid point of view	69
3.3.2.3. Methodological point of view	69
Chapter 4. Internal Correspondence Analysis	71
4.1. Running the analyses	71
4.2. Synonymous codon usage	75

4.3. Non-synonymous codon usage	75
4.3.1. Between-group analysis	75
4.3.2. Within-group analysis	80
4.3.2.1. Regular factors	80
4.3.2.2. Greedy factors	82
Conclusion	87
Appendices	89
Appendix 1	91
A1.1. Introduction	91
A1.2. Chapter 1	96
A1.3. Chapter 2	105
A1.4. Chapter 3	116
A1.5. Chapter 4	121
Appendix 2	131
A2.1. Session information	131
References	133
Index	147