Contents

1. Introduction 1

1.1 The study of relationships 1
1.2 Objectives 3
1.3 Canonical analysis: overview 4

PART I. THEORY 11

2. Canonical correlations and canonical variates 13

2.1 Introduction 13
2.2 Formulation 13
2.3 Derivation of canonical correlation coefficients and canonical variates 15
2.3.1 Eigenanalysis 15
2.3.2 Singular value decompositon 17
2.3.3 Other derivations 18
2.3.4 Concluding remarks 20
2.4 Properties of canonical correlation coefficients, weights and variates 22
2.4.1 Properties of canonical correlation coefficients 22
2.4.2 Properties of canonical weights 25
2.4.3 Properties of canonical variates 27
2.5 Computation 31
2.5.1 Numerical methods 31
2.5.2 Further remarks 35

3. Extensions and generalizations 37

3.1 Introduction 37
3.2 Further interpretive devices 37
3.2.1 Correlations between canonical variates and the original variables 38
3.2.2 Variance extracted by a canonical variate 40
3.2.3 Redundancy 40
3.2.4 Total redundancy 41

3.2.5 Variable communalities 42
3.2.6 Concluding remarks 43
3.3 Extensions and generalizations 45
3.3.1 Redundancy analysis: an alternative to canonical analysis 45
3.3.2 Improving the interpretability of canonical weights . . . 45
3.3.3 Rotation of canonical variates 50
3.3.4 Validation 50
3.3.5 Predicting a criterion of maximum utility 52
3.3.6 Generalizations of canonical analysis 53
3.3.7 Concluding remarks 55
3.4 Hypothesis testing 56
3.4.1 Independence 56
3.4.2 Dimensionality 59
3.4.3 The contribution of particular variables 62
3.4.4 Hypothesis tests for nonnormal data 63
3.4.5 Residuals from a fitted model 63

4. Canonical variate analysis 67

4.1 Introduction 67
4.2 Binary-valued dummy variables 68
4.3 Formulation and derivation 69
4.3.1 Point conceptualizations of ${}_N\mathbf{X}_p$ and ${}_N\mathbf{Z}_q$ 70
4.3.2 Derivation 73
4.4 Further aspects of canonical variate analysis 75
4.5 Hypothesis testing 80
4.5.1 Equality of g vector-means 81
4.5.2 Dimensionality 81
4.6 Affinities with other methods 82
4.6.1 Canonical variate analysis, multivariate analysis of variance and multiple discriminant analysis 82
4.6.2 Canonical variate analysis and principal component analysis 87
4.7 Imposition of structure 88
4.7.1 Designed comparisons 89
4.7.2 Separating the sources of variation 90
4.7.3 Further comments 91
4.8 Concluding remarks 92

5. Dual scaling 96

5.1 Introduction 96
5.2 Formulation and derivation 98
5.2.1 Maximizing the correlation between rows and columns 99
5.2.2 Maximizing the separation between rows and columns 103

5.3 Further aspects of dual scaling . . . 107
5.4 Hypothesis testing . . . 113
5.4.1 Independence . . . 113
5.4.2 Dimensionality . . . 114
5.5 Affinities with other methods . . . 115
5.5.1 Dual scaling and the analysis of contingency tables . . . 115
5.5.2 Dual scaling, correspondence analysis and principal component analysis . . . 119
5.6 Relationships among statistical methods . . . 121
5.7 Concluding remarks . . . 124

PART II. APPLICATIONS . . . 127

General introduction . . . 129

6. Experiment 1: an investigation of spatial variation . . . 131

6.1 Introduction . . . 131
6.2 Results . . . 132
6.2.1 The canonical correlation coefficients . . . 132
6.2.2 Independence . . . 133
6.2.3 Dimensionality . . . 133
6.2.4 The canonical variates . . . 133
6.2.5 Variable communalities . . . 140
6.3 Conclusions . . . 141

7. Experiment 2: soil-species relationships in a limestone grassland community . . . 142

7.1 Introduction . . . 142
7.2 Results . . . 143
7.2.1 The canonical correlation coefficients . . . 143
7.2.2 Independence . . . 144
7.2.3 Dimensionality . . . 144
7.2.4 The canonical variates . . . 145
7.2.5 Variable communalities . . . 150
7.3 Conclusions . . . 151

8. Soil-vegetation relationships in a lowland tropical rain forest . . . 154

8.1 Introduction . . . 154
8.2 Results . . . 155
8.2.1 The canonical correlation coefficients . . . 155
8.2.2 Independence . . . 156
8.2.3 Dimensionality . . . 156
8.2.4 The canonical variates . . . 158
8.2.5 Variable communalities . . . 162

8.3 Ecological assessment of the results . . . 162
8.4 Conclusions . . . 164

9. Dynamic status of a lowland tropical rain forest . . . 165

9.1 Introduction . . . 165
9.2 Results . . . 167
9.2.1 The canonical correlation coefficients . . . 167
9.2.2 Independence . . . 167
9.2.3 Dimensionality . . . 167
9.2.4 The canonical variates . . . 168
9.2.5 Variable communalities . . . 174
9.3 Ecological assessment of the results . . . 174
9.4 Conclusions . . . 175

10. The structure of grassland vegetation in Anglesey, North Wales . . . 177

10.1 Introduction . . . 177
10.2 Results . . . 179
10.2.1 The canonical correlation coefficients . . . 179
10.2.2 Equality of community centroids . . . 179
10.2.3 Collinearity . . . 179
10.2.4 The canonical variates . . . 180
10.2.5 Variable communalities . . . 183
10.3 Ecological assessment of the results . . . 184
10.4 Conclusions . . . 186

11. The nitrogen nutrition of eight grass species . . . 188

11.1 Introduction . . . 188
11.2 Multivariate analysis of variance . . . 189
11.2.1 Results . . . 189
11.2.2 Designed comparisons . . . 193
11.3 Canonical variate analysis . . . 198
11.3.1 Results . . . 199
11.4 Relationships between multivariate analysis of variance, discriminant analysis and canonical variate analysis . . . 203
11.5 Ecological assessment of the results . . . 204
11.6 Conclusions . . . 206

12. Herbivore-environment relationships in the Rwenzori National Park, Uganda . . . 207

12.1 Introduction . . . 207
12.2 Contingency table analysis . . . 209
12.2.1 Results . . . 210

12.3 Dual scaling 212
12.3.1 Results 213
12.4 Relationships between contingency table analysis and dual scaling 219
12.5 Ecological assessment of the results 220
12.6 Conclusions 224

PART III. APPRAISAL AND PROSPECT 227

13. Applications: assessment and conclusions 229

13.1 Introduction 229
13.2 Assessment 230
13.3 Conclusions 237

14. Research issues and future developments 238

14.1 Introduction 238
14.2 Data collection 239
14.2.1 Choice of variables 240
14.2.2 Experimental design 240
14.3 Initial data exploration 242
14.3.1 Homogeneity 242
14.3.2 Assessing joint distribution 245
14.3.3 Re-expressing variables 250
14.4 Potential data problems 253
14.4.1 Outlying or influential observations 254
14.4.2 Long-tailed distributions 256
14.4.3 Collinearity 260
14.5 Statistical assessment 266
14.5.1 Residuals in canonical analysis 266
14.5.2 Does the model fit? 267
14.5.3 Stability of results 274
14.5.4 Miscellanea 276
14.6 Concluding remarks 285

APPENDICES 289

A.1 Multivariate regression 291

A.2 Data sets used in worked applications 301

A.3 Species composition of a limestone grassland community 308

References 309
Species' index 333
Author index 335
Subject index 341