

CONTENTS

PREFACE	xiii
---------	------

Acknowledgments	xv
References	xv
About the authors	xvii

LIST OF ABBREVIATIONS	xix
-----------------------	-----

PART I: INTRODUCTION AND REVIEW

CHAPTER 1. ELEMENTARY STATISTICS BACKGROUND	2
---	---

<i>Data preview</i>	2
1.0. Introduction	4
1.1. A review of the basics	5
1.2. Statistical inference versus statistical description	10
1.2.1. The concept of estimation: points and intervals	13
1.2.2. The meaning of a confidence region	14
1.2.3. The role of the sampling distribution in statistical estimation	15
1.2.4. Large- versus small-sample theory	16
1.2.5. Hypotheses and hypothesis testing	16
1.2.6. Errors in hypothesis testing	17
1.2.7. Critical values and test statistics	18
1.3. Variable transformations	19
1.3.1. The Shapiro-Wilk statistic and symmetry	21
1.3.2. Bivariate linearity	23
1.3.3. Constant variance	26
1.3.4. Implications	32
1.4. One more viewpoint: special normality and constant variance assessments in preparation for multivariate analyses	34
1.5. A brief excursion into nonparametric statistics	35
1.6. References	39

CHAPTER 2. INFORMATION CONTENT IN GEO-REFERENCED DATA	42
---	----

2.0. Introduction	42
2.1. Elementary and spatial statistics	43
2.2. Redundant attribute and locational information	46
2.3. Exploring residuals in spatial analysis	51
2.4. References	53

Contents

CHAPTER 3. INTRODUCTION TO MATRIX ALGEBRA 55

3.0.	Elementary concepts of matrices	55
3.0.1.	Matrix operations	56
3.1.	Eigenfunctions and their meanings	58
3.2.	Illustrative eigenfunction uses in multivariate analysis	62
3.3.	Illustrative eigenfunction uses in spatial statistical analysis	66
3.4.	References	74

CHAPTER 4. MULTIPLE LINEAR REGRESSION AND CORRELATION ANALYSIS ($P \geq 3$) 79

<i>Chapter preview</i>		79
4.0.	Introduction	79
4.1.	Background	80
4.2.	Variable selection and conventional diagnostics	85
4.2.1.	Statistical distribution properties of b	87
4.2.2.	Selection criteria for the matrix of predictor variables X	89
4.2.3.	Calculation error diagnostics	97
4.2.4.	Other useful diagnostics	101
4.3.	Exploring nonlinearities: trend surface models and polynomial regression	108
4.4.	Analysis of variance (ANOVA) as a regression problem	111
4.4.1.	Basic experimental design	112
4.4.2.	N-way balanced designs	118
4.5.	Analysis of covariance ($P \geq 2, K > 3$)	126
4.6.	Effects of spatial autocorrelation	128
4.7.	Problems	130
4.8.	References	130

PART II: INSTANCES OF THE GENERAL LINEAR MODEL

CHAPTER 5. MULTIVARIATE ANALYSIS OF VARIANCE 133

<i>Chapter preview</i>		133
5.0.	Introduction	133
5.1.	N-way unbalanced designs	136
5.1.1.	A two-way unbalanced design: the fixed effects model	137
5.1.2.	A two-way unbalanced design: the mixed effects model	141
5.1.3.	Two-way ANOVA interpretation when interaction is significant	144

Contents

5.2.	Multicollinearity and simultaneous testing: Multivariate Analysis of Variance (MANOVA)	145
5.2.1.	The MANOVA format	147
5.2.2.	Selected diagnostics for MANOVA	151
5.3.	Two-way MANOVA	152
5.3.1.	Fixed effects MANOVA	152
5.3.2.	Mixed effects MANOVA	154
5.4.	Problems	156
5.5.	References	157
 CHAPTER 6. PRINCIPAL COMPONENTS AND FACTOR ANALYSIS		158
<i>Chapter preview</i>		158
6.0.	Introduction	158
6.1.	Selected mathematical features of PCA	160
6.2.	Multicollinearity	162
6.3.	Moving from PCA to factor analysis: seeking parsimony	166
6.4.	PCA/FA and hypothesis testing	167
6.5.	PCA/FA and spatial autocorrelation	171
6.6.	PCA and dimensional hypothesis testing	176
6.7.	Artificial orthogonality: principal components regression	178
6.8.	A multivariate normality diagnostic	180
6.9.	Problems	183
6.10.	References	184
 CHAPTER 7. DISCRIMINANT FUNCTION ANALYSIS		185
<i>Chapter preview</i>		185
7.0.	Introduction	185
7.1.	Properties of the linear DFA model	187
7.1.1.	Notable assumptions for the linear DFA model	187
7.1.2.	The DFA eigenfunction problem	188
7.1.3.	Significance testing in DFA	194
7.2.	Interpretation of the dimensions of regional differences	196
7.3.	Probabilities for regional membership classifications	200
7.4.	DFA as a regression problem: Two-regions DFA	202
7.5.	Problems	205
7.6.	References	205

Contents

CHAPTER 8. CLUSTER ANALYSIS 207

<i>Chapter preview</i>	207
8.0. Introduction	207
8.1. Principles of cluster analysis	208
8.2. A comparison of clustering algorithm solutions	213
8.3. Location-allocation (L-A) and cluster analysis	220
8.4. Natural groups of coterminous areal units	226
8.5. Using DFA to improve CA solutions	228
8.6. Aggregation effects	229
8.7. Problems	232
8.8. References	232

CHAPTER 9. CANONICAL CORRELATION ANALYSIS 234

<i>Chapter preview</i>	234
9.0. Introduction	235
9.1. Properties of the linear CCA model	237
9.1.1. The CCA eigenfunction problem	237
9.1.2. Mathematical features of canonical dimensions	239
9.1.3. Significance testing in CCA	245
9.2. Interpretation of the canonical dimensions spanning geo-referenced data sets	246
9.2.1. Dimensions spanning geo-referenced coordinates and areal unit attributes	250
9.2.2. Dimensions spanning geo-referenced coordinates and eigenvectors of the connectivity matrix $\mathbf{C}$	251
9.3. DFA as a CCA problem: multiple-regions DFA	254
9.4. The GLM and CCA	255
9.5. Problems	258
9.6. References	259

PART III: NONLINEAR AND CATEGORICAL DATA MODELING

CHAPTER 10. NONLINEAR REGRESSION ANALYSIS 261

<i>Chapter preview</i>	261
10.0. Introduction	261
10.1. Error structure options	262
10.2. Principles of nonlinear least squares	263

Contents

10.2.1.	Methods for defining Δ (<i>advanced material</i>)	265
10.3.	Some statistical theory for NLS estimation (<i>advanced material</i>)	267
10.4.	A comparison of various models for estimating empirical probabilities	267
10.5.	Problems	273
10.6.	References	274

CHAPTER 11. SPATIAL AUTOREGRESSIVE ANALYSIS 275

<i>Chapter preview</i>		275
11.0.	Introduction	275
11.1.	Specifying the n-by-n matrix $L_j(\rho)$	276
11.2.	Computing the normalizing constant	277
11.3.	Estimating the parameters of an SAR model	285
11.4.	Estimating the parameters of an AR model	289
11.5.	A repeated regressions interpretation of the spatial statistical analysis	294
11.6.	Problems	298
11.7.	References	299

CHAPTER 12. SPECIAL NONLINEAR REGRESSION APPLICATIONS IN SPATIAL ANALYSIS 301

<i>Chapter preview</i>		301
12.0.	Introduction	301
12.1.	The binomial distribution and logistic regression	302
12.2.	The binomial distribution and measurement error	306
12.3.	The multinomial distribution and polychotomous logistic regression	309
12.4.	A case study of children's elevated blood lead levels in Syracuse, New York	312
12.5.	Poisson regression	315
12.6.	Problems	320
12.7.	References	320

EPILOGUE 322

APPENDICES 325

A:	Puerto Rico (PR) agricultural data set	326
B:	Puerto Rico connectivity matrix, C , data set	329
C:	Samples drawn for use in examples PR-4.10 and PR-4.11	331
D:	Supplemental PR milk production data set	332

Contents

E: Puerto Rico urban population components data set	335
F: Syracuse pediatric blood lead levels data set	337
G: Metropolitan Toronto origin-destination data set	339
INDEX	340