CONTENTS

About the Author xiii
Preface to the First Edition xiv
Preface to the SecondEdition xvii
Preface to the Third Edition xix
Preface to the Fourth Edition xx
Preface to the Fifth Edition xxi
Online Resources xxii

1 INTRODUCTION TO STATISTICAL METHODS FOR GEOGRAPHY 1

1.1 Introduction 1
1.2 The scientific method 2
1.3 Exploratory and confirmatory approaches in geography 4
1.4 Probability and statistics 5
1.4.1 Probability 5
1.4.2 Statistics 6
1.4.3 Probability paradoxes 7
1.4.4 Geographical applications of probability and statistics 10
1.5 Descriptive and inferential methods 15
1.6 The nature of statistical thinking 17
1.7 Special considerations for spatial data 18
1.7.1 The modifiable areal unit problem 18
1.7.2 Boundary problems 19
1.7.3 Spatial sampling procedures 20
1.7.4 Spatial autocorrelation 20
1.8 The structure of the book 20
1.9 Datasets 22
1.9.1 Introduction 22
1.9.2 Home sales in Milwaukee, Wisconsin, USA in 2012 22

1.9.3 Singapore census data for 2010 23
1.9.4 Hypothetical UK housing prices 23
1.9.5 1990 Census Data for Erie County, New York 25
1.9.6 Monthly rain gauge accumulations for Seattle 25
1.9.7 PM2.5 Particulate matter data for Buffalo, New York (2018) 25

2 DESCRIPTIVE STATISTICS 27
2.1 Types of data 27
2.2 Visual descriptive methods 28
2.3 Measures of central tendency 32
2.4 Measures of variability 35
2.5 Other numerical measures for describing data 36
2.5.1 Coefficient of variation 36
2.5.2 Skewness 37
2.5.3 Kurtosis 38
2.5.4 Standard scores 39
2.6 Descriptive spatial statistics 39
2.6.1 The measurement of distance 39
2.6.2 Mean center 41
2.6.3 Median center 42
2.6.4 Standard distance 44
2.6.5 Relative distance 45
2.6.6 Illustration of spatial measures of central tendency and dispersion 46
2.6.7 Angular data 47
2.7 Descriptive statistics in *SPSS 25 for Windows* 51
2.7.1 Data input 51
2.7.2 Descriptive analysis 51
Solved exercises 52
Exercises 54

3 PROBABILITY AND DISCRETE PROBABILITY DISTRIBUTIONS 61
3.1 Introduction 61
3.2 Sample spaces, random variables, and probabilities 62
3.3 Binomial processes and the binomial distribution 67
3.4 The geometric distribution 71
3.5 The Poisson distribution 73
3.6 The hypergeometric distribution 80
3.6.1 Application to residential segregation 81
3.6.2 Application to the space–time clustering of disease 82

3.7 Binomial tests in *SPSS 25 for Windows* 84
Solved exercises 84
Exercises 89

4 CONTINUOUS PROBABILITY DISTRIBUTIONS AND PROBABILITY MODELS 96
4.1 Introduction 96
4.2 The uniform or rectangular distribution 97
4.3 The normal distribution 99
4.4 The exponential distribution 105
4.5 Summary of discrete and continuous distributions 111
4.6 Probability models 113
4.6.1 The intervening opportunities model 113
4.6.2 A model of migration 118
4.6.3 The future of the human population 119
Solved exercises 120
Exercises 123

5 INFERENTIAL STATISTICS: CONFIDENCE INTERVALS, HYPOTHESIS TESTING, AND SAMPLING 126
5.1 Introduction to inferential statistics 126
5.2 Confidence intervals 127
5.2.1 Confidence intervals for the mean 127
5.2.2 Confidence intervals for the mean when the sample size is small 130
5.2.3 Confidence intervals for the difference between two means 130
5.2.4 Confidence intervals for proportions 132
5.3 Hypothesis testing 133
5.3.1 Hypothesis testing and one-sample z-tests of the mean 133
5.3.2 One-sample t-tests 138
5.3.3 One-sample tests for proportions 140
5.3.4 Two-sample tests: differences in means 143
5.3.5 Two-sample tests: differences in proportions 147
5.3.6 Type II errors and statistical power 149
5.4 Distributions of the random variable and distributions of the test statistic 151
5.5 Spatial data and the implications of nonindependence 153
5.6 Further discussion of the effects of deviations from the assumptions 155
5.6.1 One-sample test of proportions: binomial distribution – assumption of constant or equal success probabilities 155

5.6.2 One-sample test of proportions: binomial distribution – assumption of independence 156
5.6.3 Two-sample difference of means test: assumption of independent observations 158
5.6.4 Two-sample difference of means test: assumption of homogeneity 160
5.7 Sampling 160
5.7.1 Spatial sampling 162
5.7.2 Sample size considerations 163
5.8 Some tests for spatial measures of central tendency and variability 166
5.9 One-sample tests of means in *SPSS 25 for Windows* 168
5.9.1 Interpretation 168
5.10 Two-sample *t*-tests in *SPSS 25 for Windows* 169
5.10.1 Data entry 169
5.10.2 Running the *t*-test 170
5.11 Two-sample *t*-tests in *Excel* 172
Solved exercises 173
Exercises 182

6 ANALYSIS OF VARIANCE 187
6.1 Introduction 187
6.1.1 A note on the use of *F*-tables 190
6.2 Illustrations 191
6.2.1 Hypothetical swimming frequency data 191
6.2.2 Diurnal variation in precipitation 193
6.3 Analysis of variance with two categories 194
6.4 Testing the assumptions 195
6.5 Consequences of failure to meet assumptions 195
6.5.1 Normality 195
6.5.2 Homoscedasticity 196
6.5.3 Independence of observations 196
6.6 The nonparametric Kruskal–Wallis test 198
6.6.1 Illustration: diurnal variation in precipitation 198
6.6.2 More on the Kruskal–Wallis test 199
6.7 The nonparametric median test 200
6.7.1 Illustration 200
6.8 Contrasts 202
6.8.1 A priori contrasts 203

6.9 One-way ANOVA in *SPSS 25 for Windows* 203
6.9.1 Data entry 203
6.9.2 Data analysis and interpretation 204
6.10 One-way ANOVA in *Excel* 206
Solved exercises 207
Exercises 209

7 CORRELATION 216
7.1 Introduction and examples of correlation 216
7.2 More illustrations 219
7.2.1 Mobility and cohort size 219
7.2.2 Statewide infant mortality rates and income 221
7.3 A significance test for *r* 223
7.3.1 Illustration 223
7.4 The correlation coefficient and sample size 224
7.5 Spearman's rank correlation coefficient 226
7.6 Additional topics 228
7.6.1 The effect of spatial dependence on significance tests for correlation coefficients 228
7.6.2 The modifiable area unit problem and spatial aggregation 230
7.7 Correlation in *SPSS 25 for Windows* 230
7.7.1 Illustration 231
7.8 Correlation in *Excel* 234
Solved exercises 234
Exercises 235

8 DATA REDUCTION: FACTOR ANALYSIS AND CLUSTER ANALYSIS 239
8.1 Introduction 239
8.2 Factor analysis and principal components analysis 240
8.2.1 Illustration: 1990 Census data for Erie County, New York 241
8.3 Cluster analysis 247
8.3.1 More on agglomerative methods 248
8.3.2 Illustration: 1990 Census data for Erie County, New York 249
8.4 Data reduction methods in *SPSS 25 for Windows* 256
8.4.1 Factor analysis 256
8.4.2 Cluster analysis 256
Exercises 257

9 INTRODUCTION TO REGRESSION ANALYSIS 261
9.1 Introduction 261
9.2 Fitting a regression line to a set of bivariate data 264
9.2.1 Illustration: income levels and consumer expenditure 267
9.3 Regression in terms of explained and unexplained sums of squares 269
9.3.1 Illustration 272
9.4 Assumptions of regression 273
9.5 Standard error of the estimate 273
9.6 Tests for β 274
9.6.1 Illustration 274
9.7 Illustration: state aid to secondary schools 275
9.8 Linear versus nonlinear models 277
9.9 Regression in *SPSS 25 for Windows* 279
9.9.1 Data input 279
9.9.2 Analysis 280
9.9.3 Options 280
9.9.4 Output 280
9.10 Regression in *Excel* 281
9.10.1 Data input 281
9.10.2 Analysis 281
Solved exercises 282
Exercises 285

10 MORE ON REGRESSION 289
10.1 Multiple regression 289
10.1.1 Multicollinearity 291
10.1.2 Interpretation of coefficients in multiple regression 291
10.1.3 Adjusted r^2 292
10.2 Misspecification error 292
10.3 Dummy variables 294
10.3.1 Dummy variable regression in a recreation planning example 297
10.4 Multiple regression illustration: species in the Galápagos Islands 299
10.4.1 Model 1: The kitchen-sink approach 300
10.4.2 Missing values 303
10.4.3 Outliers and multicollinearity 304
10.4.4 Model 2 304

10.4.5 Model 3 307
10.4.6 Model 4 309
10.5 Variable selection 310
10.6 Regression analysis on component scores 311
10.7 Categorical dependent variable 311
10.7.1 Binary response: logistic regression 312
10.8 A summary of some problems that can arise in regression analysis 316
10.9 Multiple and logistic regression in *SPSS 25 for Windows* 318
10.9.1 Multiple regression 318
10.9.2 Logistic regression 318
Exercises 322

11 INTRODUCTION TO SPATIAL PATTERNS AND SPATIAL REGRESSION 328
11.1 Introduction 328
11.2 The analysis of point patterns 329
11.2.1 Quadrat analysis 331
11.3 Geographic patterns in real data: Moran's *I* 335
11.4 Local statistics 341
11.4.1 Introduction 341
11.4.2 Local Moran statistic 341
11.5 Introduction to spatial aspects of regression 342
11.6 Spatial lag model and neighborhood-based explanatory variables 343
11.7 Spatial regression: autocorrelated errors 344
11.8 Geographically weighted regression 345
11.9 Illustration 346
11.9.1 Ordinary least squares 347
11.9.2 Spatial regression: autocorrelated errors 350
11.9.3 Geographically weighted regression 351
11.10 Finding Moran's *I* using *SPSS 25 for Windows* 351
11.11 Finding Moran's *I* using *GeoDa* 352
11.12 Spatial Regression with *GeoDa 1.4.6* 354
Exercises 356

EPILOGUE 358

ANSWERS FOR SELECTED EXERCISES 361

APPENDIX A: STATISTICAL TABLES 369

Table A.1 Random digits 369
Table A.2 Normal distribution 373
Table A.3 Student's *t*-distribution 374
Table A.4 Cumulative distribution of student's *t*-distribution 376
Table A.5 *F*-distribution 382
Table A.6 χ^2 distribution 385

APPENDIX B: MATHEMATICAL CONVENTIONS AND NOTATION 386

B.1 Mathematical conventions 386
B.2 Mathematical notation 388
B.2.1 More examples 391

Bibliography 392
Index 399