

# Contents

<b>Preface</b>	<b>xii</b>
<b>Preface to the Second Edition</b>	<b>xv</b>
<b>Preface to the Third Edition</b>	<b>xvii</b>
<b>1 Introduction to Statistical Methods for Geography</b>	<b>1</b>
1.1 Introduction	1
1.2 The scientific method	1
1.3 Exploratory and confirmatory approaches in geography	4
1.4 Probability and statistics	4
1.4.1 Probability	4
1.4.2 Statistics	5
1.4.3 Probability paradoxes	6
1.4.4 Geographical applications of probability and statistics	9
1.5 Descriptive and inferential methods	13
1.6 The nature of statistical thinking	15
1.7 Special considerations for spatial data	16
1.7.1 The modifiable areal unit problem	16
1.7.2 Boundary problems	17
1.7.3 Spatial sampling procedures	17
1.7.4 Spatial autocorrelation	18
1.8 The structure of the book	18
1.9 Datasets	19
1.9.1 Mobile phone signal strength in Erie County, New York, USA	19
1.9.2 House sales in Tyne and Wear	20
1.9.3 1990 Census Data for Erie County, New York	23
<b>2 Descriptive Statistics</b>	<b>24</b>
2.1 Types of data	24
2.2 Visual descriptive methods	25
2.3 Measures of central tendency	28
2.4 Measures of variability	30
2.5 Other numerical measures for describing data	32
2.5.1 Coefficient of variation	32
2.5.2 Skewness	32
2.5.3 Kurtosis	33
2.5.4 Standard scores	34

2.6	Descriptive spatial statistics	35
2.6.1	Mean center	35
2.6.2	Median center	36
2.6.3	Standard distance	38
2.6.4	Relative distance	38
2.6.5	Illustration of spatial measures of central tendency and dispersion	39
2.6.6	Angular data	39
2.7	Descriptive statistics in <i>SPSS 16.0 for Windows</i>	43
2.7.1	Data input	43
2.7.2	Descriptive analysis	43
	Solved exercises	44
	Exercises	46
<b>3</b>	<b>Probability and Discrete Probability Distributions</b>	<b>53</b>
3.1	Introduction	53
3.2	Sample spaces, random variables, and probabilities	53
3.3	Binomial processes and the binomial distribution	55
3.4	The geometric distribution	59
3.5	The Poisson distribution	61
3.6	The hypergeometric distribution	65
3.6.1	Application to residential segregation	67
3.6.2	Application to the space–time clustering of disease	67
3.7	Binomial tests in <i>SPSS 16.0 for Windows</i>	69
	Solved exercises	69
	Exercises	74
<b>4</b>	<b>Continuous Probability Distributions and Probability Models</b>	<b>80</b>
4.1	Introduction	80
4.2	The uniform or rectangular distribution	80
4.3	The normal distribution	83
4.4	The exponential distribution	89
4.5	Summary of discrete and continuous distributions	93
4.6	Probability models	94
4.6.1	The intervening opportunities model	95
4.6.2	A model of migration	100
4.6.3	The future of the human population	101
	Solved exercises	102
	Exercises	105

<b>5 Inferential Statistics: Confidence Intervals, Hypothesis Testing, and Sampling</b>	<b>107</b>
5.1 Introduction to inferential statistics	107
5.2 Confidence intervals	107
5.2.1 Confidence intervals for the mean	107
5.2.2 Confidence intervals for the mean when the sample size is small	110
5.2.3 Confidence interval for the difference between two means	111
5.2.4 Confidence intervals for proportions	112
5.3 Hypothesis testing	113
5.3.1 Hypothesis testing and one-sample $z$ -tests of the mean	113
5.3.2 One-sample $t$ -tests	117
5.3.3 One-sample tests for proportions	119
5.3.4 Two-sample tests: differences in means	121
5.3.5 Two-sample tests: differences in proportions	125
5.4 Distributions of the random variable and distributions of the test statistic	126
5.5 Spatial data and the implications of nonindependence	127
5.6 Further discussion of the effects of deviations from the assumptions	129
5.6.1 One-sample test of proportions: binomial distribution – assumption of constant or equal success probabilities	129
5.6.2 One-sample test of proportions: binomial distribution – assumption of independence	130
5.6.3 Two-sample difference of means test: assumption of independent observations	132
5.6.4 Two-sample difference of means test: assumption of homogeneity	134
5.7 Sampling	134
5.7.1 Spatial sampling	135
5.7.2 Sample size considerations	136
5.8 Some tests for spatial measures of central tendency and variability	139
5.9 One-sample tests of means in <i>SPSS 16.0 for Windows</i>	141
5.9.1 Interpretation	141
5.10 Two-sample $t$ -tests in <i>SPSS 16.0 for Windows</i>	142
5.10.1 Data entry	142
5.10.2 Running the $t$ -test	143
5.11 Two-sample $t$ -tests in <i>Excel</i>	143
Solved exercises	145
Exercises	153

<b>6</b>	<b>Analysis of Variance</b>	<b>157</b>
6.1	Introduction	157
6.1.1	A note on the use of <i>F</i> -tables	160
6.2	Illustrations	160
6.2.1	Hypothetical swimming frequency data	160
6.2.2	Diurnal variation in precipitation	161
6.3	Analysis of variance with two categories	163
6.4	Testing the assumptions	164
6.5	Consequences of failure to meet assumptions	164
6.6	Implications for hypothesis tests when assumptions are not met	164
6.6.1	Normality	165
6.6.2	Homoscedasticity	165
6.6.3	Independence of observations	166
6.7	The nonparametric Kruskal–Wallis test	167
6.7.1	Illustration: diurnal variation in precipitation	167
6.7.2	More on the Kruskal–Wallis test	168
6.8	The nonparametric median test	169
6.8.1	Illustration	169
6.9	Contrasts	170
6.9.1	<i>A priori</i> contrasts	172
6.10	One-way ANOVA in <i>SPSS 16.0 for Windows</i>	172
6.10.1	Data entry	172
6.10.2	Data analysis and interpretation	173
6.11	One-way ANOVA in <i>Excel</i>	174
	Solved exercises	175
	Exercises	177
<b>7</b>	<b>Correlation</b>	<b>183</b>
7.1	Introduction and examples of correlation	183
7.2	More illustrations	186
7.2.1	Mobility and cohort size	186
7.2.2	Statewide infant mortality rates and income	187
7.3	A significance test for <i>r</i>	189
7.3.1	Illustration	189
7.4	The correlation coefficient and sample size	189
7.5	Spearman’s rank correlation coefficient	191
7.6	Additional topics	192
7.6.1	The effect of spatial dependence on significance tests for correlation coefficients	192
7.6.2	The modifiable area unit problem and spatial aggregation	194

7.7	Correlation in <i>SPSS 16.0 for Windows</i>	195
7.7.1	Illustration	195
7.8	Correlation in <i>Excel</i>	197
	Solved exercises	197
	Exercises	198
<b>8</b>	<b>Introduction to Regression Analysis</b>	<b>201</b>
8.1	Introduction	201
8.2	Fitting a regression line to a set of bivariate data	204
8.2.1	Illustration: income levels and consumer expenditure	206
8.3	Regression in terms of explained and unexplained sums of squares	207
8.3.1	Illustration	211
8.4	Assumptions of regression	211
8.5	Standard error of the estimate	212
8.6	Tests for beta	212
8.6.1	Illustration	212
8.7	Illustration: state aid to secondary schools	213
8.8	Linear versus nonlinear models	215
8.9	Regression in <i>SPSS 16.0 for Windows</i>	217
8.9.1	Data input	217
8.9.2	Analysis	217
8.9.3	Options	218
8.9.4	Output	218
8.10	Regression in <i>Excel</i>	218
8.10.1	Data input	218
8.10.2	Analysis	218
	Solved exercises	220
	Exercises	222
<b>9</b>	<b>More on Regression</b>	<b>225</b>
9.1	Multiple regression	225
9.1.1	Multicollinearity	226
9.1.2	Interpretation of coefficients in multiple regression	227
9.2	Misspecification error	227
9.3	Dummy variables	229
9.3.1	Dummy variable regression in a recreation planning example	231
9.4	Multiple regression illustration: species in the Galapagos Islands	233
9.4.1	Model 1: The kitchen-sink approach	233
9.4.2	Missing values	235

9.4.3	Outliers and multicollinearity	237
9.4.4	Model 2	238
9.4.5	Model 3	238
9.4.6	Model 4	241
9.5	Variable selection	241
9.6	Categorical dependent variable	242
9.6.1	Binary response	243
9.7	A summary of some problems that can arise in regression analysis	247
9.8	Multiple and logistic regression in <i>SPSS 16.0 for Windows</i>	247
9.8.1	Multiple regression	247
9.8.2	Logistic regression	249
	Exercises	252
<b>10</b>	<b>Spatial Patterns</b>	<b>257</b>
10.1	Introduction	257
10.2	The analysis of point patterns	258
10.2.1	Quadrat analysis	259
10.2.2	Nearest neighbor analysis	263
10.3	Geographic patterns in areal data	266
10.3.1	An example using a chi-square test	266
10.3.2	Moran's $I$	268
10.4	Local statistics	274
10.4.1	Introduction	274
10.4.2	Local Moran statistic	275
10.4.3	Getis' $G_i^*$ statistic	275
10.5	Finding Moran's $I$ using <i>SPSS 16.0 for Windows</i>	276
10.6	Finding Moran's $I$ using <i>GeoDa</i>	277
	Exercises	278
<b>11</b>	<b>Some Spatial Aspects of Regression Analysis</b>	<b>281</b>
11.1	Introduction	281
11.2	Added-variable plots	282
11.3	Spatial regression: autocorrelated errors	283
11.4	Spatially varying parameters	284
11.4.1	The expansion method	284
11.4.2	Geographically weighted regression	285
11.5	Illustration	286
11.5.1	Ordinary least squares	287
11.5.2	Added-variable plots	288
11.5.3	Spatial regression: autocorrelated errors	290

11.5.4	Expansion method	291
11.5.5	Geographically weighted regression	293
11.6	Spatial Regression with <i>GeoDa 0.9.5-i</i>	293
	Exercises	295
<b>12</b>	<b>Data Reduction: Factor Analysis and Cluster Analysis</b>	<b>297</b>
12.1	Introduction	297
12.2	Factor analysis and principal components analysis	297
12.2.1	Illustration: 1990 Census data for Erie County, New York	298
12.2.2	Regression analysis on component scores	302
12.3	Cluster analysis	303
12.3.1	More on agglomerative methods	306
12.3.2	Illustration: 1990 Census data for Erie County, New York	306
12.4	Data reduction methods in <i>SPSS 16.0 for Windows</i>	311
12.4.1	Factor analysis	311
12.4.2	Cluster analysis	311
	Exercises	313
	<b>Epilogue</b>	<b>316</b>
	<b>Appendix A: Statistical Tables</b>	<b>318</b>
Table A.1	Random digits	318
Table A.2	Normal distribution	320
Table A.3	Student's <i>t</i> -distribution	321
Table A.4	Cumulative distribution of Student's <i>t</i> -distribution	322
Table A.5	<i>F</i> -distribution	324
Table A.6	$\chi^2$ distribution	327
	<b>Appendix B: Mathematical Conventions and Notation</b>	<b>328</b>
B.1	Mathematical conventions	328
B.2	Mathematical notation	330
	<b>Appendix C: Review and Extension of Some Probability Theory</b>	<b>334</b>
C.1	Expected values	334
C.2	Variance of a random variable	336
C.3	Covariance of random variables	336
	<b>Bibliography</b>	<b>338</b>
	<b>Index</b>	<b>343</b>