

STATISTICAL
METHODS FOR
GEOGRAPHY

PETER A. ROGERSON

Contents

Preface	x
1 Introduction to Statistical Analysis in Geography	1
1.1 Introduction	1
1.2 The scientific method	1
1.3 Exploratory and confirmatory approaches in geography	4
1.4 Descriptive and inferential methods	5
1.4.1 Overview of descriptive analysis	5
1.4.2 Overview of inferential analysis	9
1.5 The Nature of statistical thinking	12
1.6 Some special considerations with spatial data	13
1.6.1 Modifiable areal unit problem	13
1.6.2 Boundary problems	14
1.6.3 Spatial sampling procedures	14
1.6.4 Spatial autocorrelation	15
1.7 Descriptive statistics in <i>SPSS for Windows 9.0</i>	15
1.7.1 Data input	15
1.7.2 Descriptive analysis	15
Exercises	16
2 Probability and Probability Models	18
2.1 Mathematical conventions and notation	18
2.1.1 Mathematical conventions	18
2.1.2 Mathematical notation	20
2.1.3 Examples	23
2.2 Sample spaces, random variables, and probabilities	23
2.3 The binomial distribution	25
2.4 The normal distribution	27
2.5 Confidence intervals for the mean	30
2.6 Probability models	31
2.6.1 The intervening opportunities model	32
2.6.2 A model of migration	36
2.6.3 The future of the human population	38
Exercises	39
3 Hypothesis Testing and Sampling	42
3.1 Hypothesis testing and one-sample z-tests of the mean	42
3.2 One-sample t-tests	46
3.2.1 Illustration	47

3.3	One-sample tests for proportions	48
3.3.1	Illustration	48
3.4	Two-sample tests	49
3.4.1	Two-sample <i>t</i> -tests for the mean	49
3.4.2	Two-sample tests for proportions	53
3.5	Distributions of the variable and the test statistic	54
3.6	Spatial data and the implications of nonindependence	55
3.7	Sampling	57
3.7.1	Spatial sampling	58
3.8	Two-sample <i>t</i> -tests in <i>SPSS for Windows 9.0</i>	59
3.8.1	Data entry	59
3.8.2	Running the <i>t</i> -test	60
	Exercises	62
4	Analysis of Variance	65
4.1	Introduction	65
4.1.1	A note on the use of <i>F</i> -tables	67
4.1.2	More on sums of squares	67
4.2	Illustrations	68
4.2.1	Hypothetical swimming frequency data	68
4.2.2	Diurnal variation in precipitation	69
4.3	Analysis of variance with two categories	70
4.4	Testing the assumptions	70
4.5	The nonparametric Kruskal–Wallis test	70
4.5.1	Illustration: diurnal variation in precipitation	71
4.5.2	More on the Kruskal–Wallis test	72
4.6	Contrasts	73
4.6.1	<i>A priori</i> contrasts	75
4.7	Spatial dependence	75
4.8	One-way ANOVA in <i>SPSS for Windows 9.0</i>	76
4.8.1	Data entry	76
4.8.2	Data analysis and interpretation	77
4.8.3	Levene's test for equality of variances	79
4.8.4	Tests of normality: the Shapiro–Wilk test	80
	Exercises	81
5	Correlation	86
5.1	Introduction and examples of correlation	86
5.2	More illustrations	89
5.2.1	Mobility and cohort size	89
5.2.2	Statewide infant mortality rates and income	91
5.3	A significance test for <i>r</i>	92
5.3.1	Illustration	93
5.4	The correlation coefficient and sample size	93
5.5	Spearman's rank correlation coefficient	94
5.6	Additional topics	96
5.6.1	Confidence intervals for correlation coefficients	96

5.6.2	Differences in correlation coefficients	97
5.6.3	The effect of spatial dependence on significance tests for correlation coefficients	97
5.6.4	Modifiable area unit problem and spatial aggregation	99
5.7	Correlation in <i>SPSS for Windows 9.0</i>	100
5.7.1	Illustration	101
	Exercises	102
6	Introduction to regression analysis	104
6.1	Introduction	104
6.2	Fitting a regression line to a set of bivariate data	107
6.3	Regression in terms of explained and unexplained sums of squares	109
6.4	Assumptions of regression	112
6.5	Standard error of the estimate	112
6.6	Tests for beta	112
6.7	Confidence intervals	113
6.8	Illustration: income levels and consumer expenditures	113
6.9	Illustration: state aid to secondary schools	116
6.10	Linear versus nonlinear models	118
6.11	Regression in <i>SPSS for Windows 9.0</i>	120
6.11.1	Data input	120
6.11.2	Analysis	120
6.11.3	Options	121
6.11.4	Output	122
	Exercises	122
7	More on Regression	124
7.1	Multiple regression	124
7.1.1	Multicollinearity	125
7.1.2	Interpretation of coefficients in multiple regression	126
7.2	Misspecification error	126
7.3	Dummy variables	128
7.3.1	Dummy variable regression in a recreation planning example	130
7.4	Multiple regression illustration: species in the Galapagos Islands	132
7.4.1	Model 1: the kitchen-sink approach	132
7.4.2	Missing values	134
7.4.3	Outliers and multicollinearity	136
7.4.4	Model 2	136
7.4.5	Model 3	138
7.4.6	Model 4	139
7.5	Variable selection	140
7.6	Categorical dependent variable	140
7.6.1	Binary response	141
7.7	A Summary of some problems that can arise in regression analysis	145

7.8 Multiple and logistic regression in <i>SPSS for Windows 9.0</i>	145
7.8.1 Multiple regression	145
7.8.2 Logistic regression	145
Exercises	150
8 Spatial Patterns	154
8.1 Introduction	154
8.2 The analysis of point patterns	154
8.2.1 Quadrat analysis	156
8.2.2 Nearest neighbor analysis	161
8.3 Geographic patterns in areal data	164
8.3.1 An example using a chi-square test	164
8.3.2 The join-count statistic	165
8.3.3 Moran's I	167
8.4 Local statistics	173
8.4.1 Introduction	173
8.4.2 Local Moran statistic	173
8.4.3 Getis's G_i^* statistic	174
8.5 Finding Moran's I Using <i>SPSS for Windows 9.0</i>	175
Exercises	176
9 Some Spatial Aspects of Regression Analysis	179
9.1 Introduction	179
9.2 Added-variable plots	180
9.3 Spatial regression	181
9.4 Spatially varying parameters	182
9.4.1 The expansion method	182
9.4.2 Geographically weighted regression	183
9.5 Illustration	184
9.5.1 Ordinary least-squares regression	186
9.5.2 Added-variable plots	186
9.5.3 Spatial regression	187
9.5.4 Expansion method	188
9.5.5 Geographically weighted regression	190
Exercises	190
10 Data Reduction: Factor Analysis and Cluster Analysis	192
10.1 Factor analysis and principal components analysis	192
10.1.1 Illustration: 1990 census data for Buffalo, New York	193
10.1.2 Regression analysis on component scores	197
10.2 Cluster analysis	197
10.2.1 More on agglomerative methods	201
10.2.2 Illustration: 1990 census data for Erie County, New York	201
10.3 Data reduction methods in <i>SPSS for Windows 9.0</i>	207
10.3.1 Factor analysis	207
10.3.2 Cluster analysis	207
Exercises	208

Epilogue	210
Selected publications	211
Appendix A: Statistical tables	212
Table A.1 Random digits	212
Table A.2 Normal distribution	214
Table A.3 Student's t distribution	215
Table A.4 Cumulative distribution of Students t distribution	216
Table A.5 F distribution	218
Table A.6 χ^2 distribution	221
Table A.7 Coefficients for the Shapiro–Wilk W Test	222
Table A.8 Critical values for the Shapiro–Wilk W Test	224
Appendix B: Review and extension of some probability theory	225
Expected values	225
Variance of a random variable	227
Covariance of random variables	227
Bibliography	229
Index	233