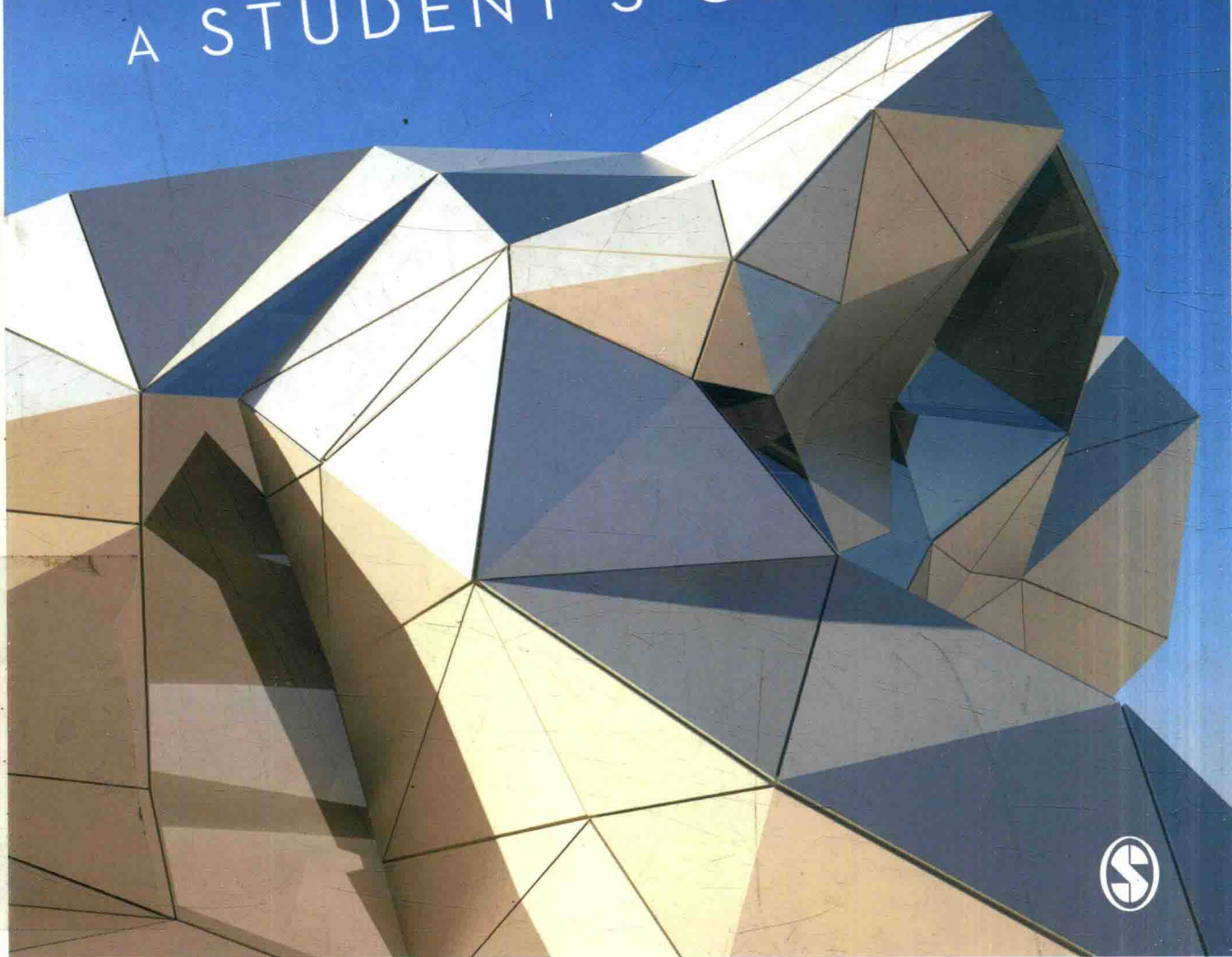


Peter A. Rogerson

FOURTH
EDITION

STATISTICAL METHODS *for* GEOGRAPHY

A STUDENT'S GUIDE



Peter A. Rogerson

STATISTICAL METHODS *for* GEOGRAPHY



A STUDENT'S GUIDE



Los Angeles | London | New Delhi
Singapore | Washington DC



Los Angeles | London | New Delhi
Singapore | Washington DC

SAGE Publications Ltd
1 Oliver's Yard
55 City Road
London EC1Y 1SP

SAGE Publications Inc.
2455 Teller Road
Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road
New Delhi 110 044

SAGE Publications Asia-Pacific Pte Ltd
3 Church Street
#10-04 Samsung Hub
Singapore 049483

Editor: Robert Rojek
Assistant editor: Keri Dickens
Production editor: Katherine Haw
Copyeditor: Catja Pafort
Proofreader: Neil Dowden
Marketing manager: Michael Ainsley
Cover design: Francis Kenney
Typeset by: C&M Digital (P) Ltd, Chennai, India
Printed in India at Replika Press Pvt Ltd



© Peter A. Rogerson 2015

This edition first published 2015
2nd edition: Reprinted 2008 and 2009, 3rd edition: Reprinted
2010, 2011 and 2013

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act, 1988, this publication may be reproduced, stored or transmitted in any form, or by any means, only with the prior permission in writing of the publishers, or in the case of reprographic reproduction, in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

Library of Congress Control Number: 2014931420

British Library Cataloguing in Publication data

A catalogue record for this book is available from
the British Library

ISBN 978-1-4462-9572-4
ISBN 978-1-4462-9573-1 (pbk)

STATISTICAL METHODS *for* GEOGRAPHY



SAGE was founded in 1965 by Sara Miller McCune to support the dissemination of usable knowledge by publishing innovative and high-quality research and teaching content. Today, we publish more than 750 journals, including those of more than 300 learned societies, more than 800 new books per year, and a growing range of library products including archives, data, case studies, reports, conference highlights, and video. SAGE remains majority-owned by our founder, and on her passing will become owned by a charitable trust that secures our continued independence.

Los Angeles | London | Washington DC | New Delhi | Singapore

此为试读, 需要完整PDF请访问: www.ertongbook.com

ABOUT THE AUTHOR

Peter A. Rogerson is SUNY (State University of New York) Distinguished Professor in the Department of Geography at the University at Buffalo, Buffalo, New York, USA. He also holds an adjunct appointment in the Department of Biostatistics and is a member of the National Center for Geographic Information and Analysis.

PREFACE

The development of geographic information systems (GIS), an increasing availability of spatial data, and recent advances in methodological techniques have all combined to make this an exciting time to study geographic problems. During the late 1970s and throughout the 1980s, there had been among many an increasing disappointment in, and questioning of, the methods developed during the quantitative revolution of the 1950s and 1960s. Perhaps this reflected expectations that were initially too high – many had thought that sheer computing power coupled with sophisticated modeling would ‘solve’ many of the social problems faced by urban and rural regions. But the poor performance of spatial analysis that was perceived by many was at least partly attributable to a limited capability to access, display, and analyze geographic data. During the last decade, geographic information systems have been instrumental in not only providing us with the capability to store and display information, but also in encouraging the provision of spatial datasets and the development of appropriate methods of quantitative analysis. Indeed, the GIS revolution has served to make us aware of the critical importance of spatial analysis. Geographic information systems do not realize their full potential without the ability to carry out methods of statistical and spatial analysis, and an appreciation of this dependence has helped to bring about a renaissance in the field.

Significant advances in quantitative geography have been made during the past decade, and geographers now have both the tools and the methods to make valuable contributions to fields as diverse as medicine, criminal justice, and the environment. These capabilities have been recognized by those in other fields, and geographers are now routinely called upon as members of interdisciplinary teams studying complex problems. Improvements in computer technology and computation have led quantitative geography in new directions. For example, the new field of geocomputation (see, e.g., Longley et al. 1998) lies at the intersection of computer science, geography, information science, mathematics, and statistics. The recent book by Fotheringham et al. (2000) also summarizes many of the new research frontiers in quantitative geography.

The purpose of this book is to provide undergraduate and beginning graduate students with the background and foundation that is necessary to be prepared for spatial analysis in this new era. I have deliberately adopted a fairly traditional approach to statistical analysis,

along with several notable differences. First, I have attempted to condense much of the material found in the beginning of introductory texts on the subject. This has been done so that there is an opportunity to progress further in important areas, such as regression analysis and the analysis of geographic patterns in one semester's time. Regression is by far the most common method used in geographic analysis, and it is unfortunate that it is often left to be covered hurriedly in the last week or two of a 'Statistics in Geography' course.

The level of the material is aimed at upper-level undergraduate and beginning graduate students. I have attempted to structure the book so that it may be used as either a first-semester or second-semester text. It may be used for a second-semester course by those students who already possess some background in introductory statistical concepts. The introductory material here would then serve as a review. However, the book is also meant to be fairly self-contained, and thus it should also be appropriate for those students learning about statistics in geography for the first time. First-semester students, after completing the introductory material in the first few chapters, will still be able to learn about the methods used most often by geographers by the end of a one-semester course; this is often not possible with many first-semester texts.

In writing this text, I had several goals. The first was to provide the basic material associated with the statistical methods most often used by geographers. Since a very large number of textbooks provide this basic information, I also sought to distinguish it in several ways. I have attempted to provide plenty of exercises. Some of these are to be done by hand (in the belief that it is always a good learning experience to carry out a few exercises by hand, despite what may sometimes be seen as drudgery!), and some require a computer. Although teaching the reader how to use computer software for statistical analysis is *not* one of the specific aims of this book, some guidance on the use of *SPSS for Windows* is provided. It is important that students become familiar with *some* software that is capable of statistical analysis. An important skill is the ability to sift through output and pick out what is important and what is not. Different software will produce output in different forms, and it is also important to be able to pick out relevant information whatever the arrangement of output.

In addition, I have tried to give students some appreciation for the special issues and problems raised by the use of geographic data. Straightforward application of the standard methods ignores the special nature of spatial data, and can lead to misleading results. Topics such as spatial autocorrelation and the modifiable areal unit problem are introduced to provide a good awareness of these issues, their consequences, and potential solutions. Because a full treatment of these topics would require a higher level of mathematical sophistication, they are not covered fully, but pointers to other, more advanced work and to examples are provided.

Another objective has been to provide some examples of statistical analysis that appear in the recent literature in geography. This should help to make clear the relevance and timeliness of the methods. Finally, I have attempted to point out some of the limitations of

a confirmatory statistical perspective, and have directed the student to some of the newer literature on exploratory spatial data analysis. Despite the popularity and importance of exploratory methods, inferential statistical methods remain absolutely essential in the assessment of hypotheses. This text aims to provide a background in these statistical methods and to illustrate the special nature of geographic data.

A Guggenheim Fellowship afforded me the opportunity to finish the manuscript during a sabbatical leave in England. I would like to thank Paul Longley for his careful reading of an earlier draft of the book. His excellent suggestions for revision have led to a better final result. Yifei Sun and Ge Lin also provided comments that were very helpful in revising earlier drafts. Art Getis, Stewart Fotheringham, Chris Brunsdon, Martin Charlton, and Ikuho Yamada suggested changes in particular sections, and I am grateful for their assistance. Emil Boasson and my daughter, Bethany Rogerson, assisted with the production of the figures. I am thankful for the thorough job carried out by Richard Cook in editing the manuscript. Finally, I would like to thank Robert Rojek at SAGE Publications for his encouragement and guidance.

PREFACE TO THE SECOND EDITION

In the first edition of *Statistical Methods for Geography*, a primary goal was to ensure adequate introductory treatment of regression and the statistical analysis of geographic patterns. These topics are of paramount importance to geographers, and there often is not time in a first course on quantitative methods to give them the attention that they deserve. A primary goal in this second edition has been to give more space to introductory topics in probability and statistics; consequently, the presentation of this important foundational material is not as condensed as it was in the previous edition. As a result, I am hopeful that the book will now be more appealing as a choice for a first-semester, as well as a second-semester text.

In particular, more introductory material has been added to the first chapter. Much of the second chapter, on descriptive statistics, is entirely new. Material on additional probability distributions has been added to what are now Chapters 3 and 4. In addition, the number of exercises has been increased, and implementation of the methods is described for users of *Excel*, as well as for users of *SPSS*. Some of the added exercises are built around two new datasets – one consisting of measures of cell or mobile phone signal strength for over 200 geographic locations near Buffalo, New York, and the other consisting of the sales price for over 500 houses, along with associated characteristics of both the houses and their neighborhoods in Tyne and Wear.

Some of the more technical and peripheral material from the first edition has been removed, and greater emphasis has been placed upon interpretation. For example, operational details for statistical tests of the assumptions of homoscedasticity and normality have been dropped, and there is now a more extended discussion of the consequences of the failure to meet the assumptions of the various tests. In particular, further emphasis has been placed upon the assumption of statistical dependence.

The development of this second edition has benefited from the many constructive suggestions made by the anonymous reviewers who were solicited to provide feedback on the first edition and ideas for the second. Barry Solomon and others pointed out typos in the first edition, and I acknowledge gratefully their contribution. Jessie Poon pointed out a

number of recent and relevant articles. Daejoing Kim assisted with the production of several figures, and Gyoungju Lee provided helpful comments. I thank Stewart Fotheringham for providing the Tyne and Wear housing data. I am grateful to Vanessa Harwood at Sage for her fine efforts in the production of the manuscript. I also thank Anjana Narayanan and Frances Morgan at Keyword Group for their work with the manuscript. Finally, I would once again like to thank Robert Rojek at SAGE for his efforts and advice. It is not uncommon for the life cycle of a book to experience several different editors (which says something about the time it takes to develop a book and/or the turnover rate of editors!); I have been fortunate to have the steady guidance of a top-notch editor for both editions.

PREFACE TO THE THIRD EDITION

Two of the primary goals in this third edition have been to (a) add illustrative, worked exercises at the end of each chapter, and (b) give additional attention to the implementation of spatial regression (through the use of the free and widely used *GeoDa* software; see Anselin et al. 2006). With regard to the former, the worked exercises are intended to supplement both the illustrations that already appear within the text and the exercises that appear at the end of each chapter. They are what might be regarded as ‘mixed exercises’; unlike the examples that are introduced with each concept, these require the reader to think about the collection of ideas introduced in the chapter. With regard to the latter, it is essential that geographers be aware of the effects of spatial dependence on statistical analyses, and that they be able to use tools designed to uncover geographic patterns, and additions to this edition should help in that regard.

The third edition has benefited from constructive suggestions made by the anonymous reviewers who were solicited by SAGE to provide feedback on the second edition and ideas for the third. I would like to thank Jared Aldstadt, Greg Aleksandrov, Richard Brandt, Jerry Davis, David Mark, Tom Pancykowski, Gregory Taff, and Enki Yoo for pointing out typos and inconsistencies in the second edition; and I would like to acknowledge the contribution to this edition made by Raymond Greene – Department of Geography, Western Illinois University – who prepared the chapter ancillaries on the book’s companion website (www.sagepub.co.uk/rogerson). I am grateful to Sarah-Jayne Boyd at SAGE for her assistance with the manuscript. Finally, I would like to thank Katherine Haw, and Robert Rojek at SAGE a third time for all of his kind guidance.

PREFACE TO THE FOURTH EDITION

This new edition includes new sections on the measurement of distance (in Chapter 2) and Type II errors and the concept of statistical power (Chapter 5). Perhaps more importantly, a complete re-working of the presentations, explanations, and examples from the third edition should (I hope!) improve both the clarity and the reader's level of understanding.

Another new feature is the addition of 'Answers for Selected Exercises', which appear at the end of the book.

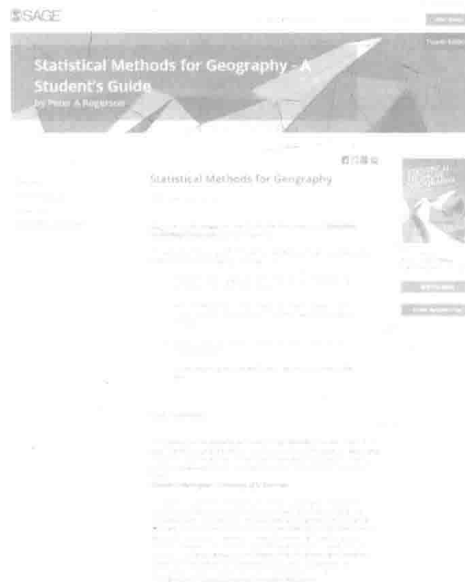
Like the previous editions, this one has also benefited from the helpful comments from many people. I am grateful to Jared Aldstadt, Richard Brandt, Sarah Glann, Eva Johansson, Paul Johnson, Xiaoxing Qin, Fernando Rios, and Enki Yoo for pointing out typos and for their suggestions. Rusty Weaver assisted by providing information that was useful in updating some of the material. I am also grateful for Peter Kedron's assistance with a number of the figures. I would especially like to thank Keri Dickens and Katherine Haw at SAGE for their help in putting this edition together. Finally, I am grateful for the support of Robert Rojek at SAGE – he has now provided both encouragement and sound advice on all four editions.

ABOUT THE COMPANION WEBSITE

PRELIMINARY MATERIAL: GUIDE TO HOW TO USE THIS WEBSITE

Specially developed for the fourth edition, be sure to visit the *Statistical Methods for Geography* companion website at **study.sagepub.com/rogerson4e** to find a range of interactive teaching and learning materials for both lecturers and students including:

- Datasets: Each dataset utilised in the text is included with a description of its format and fields
- Further Resources: Links to websites, videos, podcasts, and other media which supplements the information presented in the text
- Hyperlinked text that provides further information on the topics discussed
- Links to journal articles and book chapters mentioned in the text



CONTENTS

<i>About the Author</i>	xiii
<i>Preface</i>	xiv
<i>Preface to the Second Edition</i>	xvii
<i>Preface to the Third Edition</i>	xix
<i>Preface to the Fourth Edition</i>	xx
<i>About the Companion Website</i>	xxi

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	6
1.4.4 Geographical applications of probability and statistics	10
1.5 Descriptive and inferential methods	14
1.6 The nature of statistical thinking	16
1.7 Special considerations for spatial data	17
1.7.1 The modifiable areal unit problem	17
1.7.2 Boundary problems	18
1.7.3 Spatial sampling procedures	19
1.7.4 Spatial autocorrelation	19
1.8 The structure of the book	19
1.9 Datasets	21
1.9.1 Introduction	21
1.9.2 Home sales in Milwaukee, Wisconsin, USA in 2012	21
1.9.3 Singapore census data for 2010	22

1.9.4	Hypothetical UK housing prices	22
1.9.5	1990 Census Data for Erie County, New York	24
2	DESCRIPTIVE STATISTICS	26
2.1	Types of data	26
2.2	Visual descriptive methods	27
2.3	Measures of central tendency	31
2.4	Measures of variability	33
2.5	Other numerical measures for describing data	35
2.5.1	Coefficient of variation	35
2.5.2	Skewness	35
2.5.3	Kurtosis	36
2.5.4	Standard scores	37
2.6	Descriptive spatial statistics	38
2.6.1	The measurement of distance	38
2.6.2	Mean center	40
2.6.3	Median center	41
2.6.4	Standard distance	43
2.6.5	Relative distance	44
2.6.6	Illustration of spatial measures of central tendency and dispersion	44
2.6.7	Angular data	46
2.7	Descriptive statistics in <i>SPSS 21 for Windows</i>	49
2.7.1	Data input	49
2.7.2	Descriptive analysis	49
	Solved exercises	51
	Exercises	53
3	PROBABILITY AND DISCRETE PROBABILITY DISTRIBUTIONS	60
3.1	Introduction	60
3.2	Sample spaces, random variables, and probabilities	61
3.3	Binomial processes and the binomial distribution	63
3.4	The geometric distribution	67
3.5	The Poisson distribution	70
3.6	The hypergeometric distribution	74
3.6.1	Application to residential segregation	76
3.6.2	Application to the space-time clustering of disease	76
3.7	Binomial tests in <i>SPSS 21 for Windows</i>	78
	Solved exercises	79
	Exercises	83

4 CONTINUOUS PROBABILITY DISTRIBUTIONS AND PROBABILITY MODELS	90
4.1 Introduction	90
4.2 The uniform or rectangular distribution	91
4.3 The normal distribution	93
4.4 The exponential distribution	99
4.5 Summary of discrete and continuous distributions	104
4.6 Probability models	106
4.6.1 The intervening opportunities model	106
4.6.2 A model of migration	111
4.6.3 The future of the human population	112
Solved exercises	113
Exercises	116
5 INFERENCE STATISTICS: CONFIDENCE INTERVALS, HYPOTHESIS TESTING, AND SAMPLING	119
5.1 Introduction to inferential statistics	119
5.2 Confidence intervals	120
5.2.1 Confidence intervals for the mean	120
5.2.2 Confidence intervals for the mean when the sample size is small	122
5.2.3 Confidence intervals for the difference between two means	123
5.2.4 Confidence intervals for proportions	125
5.3 Hypothesis testing	126
5.3.1 Hypothesis testing and one-sample z -tests of the mean	126
5.3.2 One-sample t -tests	130
5.3.3 One-sample tests for proportions	133
5.3.4 Two-sample tests: differences in means	135
5.3.5 Two-sample tests: differences in proportions	139
5.3.6 Type II errors and statistical power	141
5.4 Distributions of the random variable and distributions of the test statistic	142
5.5 Spatial data and the implications of nonindependence	144
5.6 Further discussion of the effects of deviations from the assumptions	146
5.6.1 One-sample test of proportions: binomial distribution – assumption of constant or equal success probabilities	146
5.6.2 One-sample test of proportions: binomial distribution – assumption of independence	147
5.6.3 Two-sample difference of means test: assumption of independent observations	149
5.6.4 Two-sample difference of means test: assumption of homogeneity	151