An Introduction to

STATISTICAL

PROBLEM SOLVING IN

GEOGRAPHY

second edition

J. Chapman McGrew, Jr.

Charles B. Monroe

AN INTRODUCTION TO STATISTICAL PROBLEM SOLVING IN GEOGRAPHY

SECOND EDITION

AN INTRODUCTION TO STATISTICAL PROBLEM SOLVING IN GEOGRAPHY

SECOND EDITION

J. Chapman McGrew, Jr.
Salisbury State University

Charles B. Monroe
The University of Akron

*This text is dedicated to our patient and tolerant wives,
Kathy and Laura. The innumerable hours spent working on this
manuscript meant less time could be devoted to family activities
and to our children, Courtney, Grady, Caroline, and Michael. The
continuous unselfish support of this project by our families is
deeply appreciated.*

McGraw-Hill Higher Education ✍

A Division of The McGraw-Hill Companies

This book is printed on acid-free paper.

www.mhhe.com

PREFACE

All well-trained geographers need to be proficient in applying statistical techniques to the problems they face. Geography students need a sound introduction to the variety of ways in which statistical procedures are used to explore and solve practical geographic problems. The second edition of this book continues under the principles of the first edition and is designed to provide a comprehensive and understandable introduction to statistical methods in a problem-solving framework. We want students who use this textbook to become well-grounded in and feel comfortable applying statistical techniques in situations they will encounter in their subsequent geographic education and career.

This book is targeted for undergraduate geography majors and entry-level graduate students with limited backgrounds in statistical approaches to geographic problem solving. Because this is an introductory textbook, we assume that students have not taken other courses in statistical analysis and do not have previous experience with statistical methods. However, students with a background in statistics will still find the problem-solving emphasis useful.

In this second edition, many of the features of the first edition are continued and, in many areas, developed further. In this edition, we continue to stress the importance of written narratives that explain each statistical technique fully in ways that undergraduate geography majors and beginning graduate students can easily understand. This is done without compromising or oversimplifying the statistical integrity of the material. As in the first edition, real-world examples and problems are drawn from a variety of topical areas in both human and physical geography, and these examples are fully integrated into the text. Problems are again taken from various spatial levels that range from local to international.

The use of spatial statistics, both for descriptive and inferential purposes, continues to be emphasized. We carry on the practices begun in the first edition of using a flexible, exploratory approach that emphasizes real-world problem solving and the p-value method of statistical testing.

Instructors who have used the first edition of *An Introduction to Statistical Problem Solving in Geography* should note these major changes that have been incorporated into the second edition:

- Throughout the book, geography receives increased emphasis. For example, as a technique is introduced, a more detailed, student-friendly discussion about the geographic applications of the technique is included, with more written narrative examples. This change does not mean a decreased role of math and statistics, but rather more supportive narrative material throughout the text to help the geography student master beginning statistical techniques.
- Many new examples are presented throughout the text, particularly in the chapters dealing with inferential techniques. Most of the example problems found in chapters 9, 10, and 11 of the first edition have been replaced by new examples or have been revised. A new geographic example has been added in the second edition for each of the following techniques: two-sample difference of means Z or t test, two-sample difference of proportions test, matched-pairs (dependent-sample) difference test, ANOVA and Kruskal-Wallis test, goodness-of-fit chi-square uniform, goodness-of-fit chi-square proportional, goodness-of-fit Kolmogorov-Smirnov test for normality, contingency analysis, and area pattern analysis.
- A student-directed list of "major goals and objectives" has been added at the end of each chapter.

These goals and objectives will allow students to monitor their own progress and their mastery of geographic statistical materials.

- A new epilogue chapter has been added. This chapter contains a set of more than 150 "geographic situations." The ultimate goal of spatial analysis and the use of statistical techniques in geography is to be able to determine what statistical technique should be used in a given geographic situation. Instructors will find this feature useful to assess student abilities at the completion of the course.

Other narrative and explanatory materials are new to the second edition, including the following:

- The discussion in chapter 1 concerning how geographers use statistics, particularly in the context of a general research process, has been expanded. A new example discussing acid deposition across North America has been added, providing another demonstration of how geographers use statistics in the research process.
- In chapter 5, a new section presents "Basic Probability Terms and Concepts," including a detailed introduction of basic probability rules, laws, and formulas that guide the use of probability. This is accompanied by geographic examples to illustrate the principles discussed. Covered in greater depth are such concepts as statistical independence, complements, the simple and extended multiplication rule of probability, and the addition rules for mutually exclusive and nonmutually exclusive events.
- The material in chapter 7 has been reorganized, and new material has been included. These changes are designed to provide smoother sequencing and more student-friendly explanations of such difficult concepts as the central limit theorem, standard error of the mean, and the finite correction factor. Also, the confidence interval equations associated with tables 7.3 and 7.4 in the first edition have been reorganized and fragmented into smaller tables that are easier for the student to follow.
- The material in chapter 8 has been reorganized to provide a smoother, more logical flow. The one-sample difference of means tests are now separated from the one-sample difference of proportions test. Also, a new table supplies a schematic framework to classify statistical methods. Rows of this table represent different levels of measurement, and the columns group statistical approaches according to the type of question being asked.
- The material grouped together in chapter 9 of the first edition has now been split into two distinct chapters: "Two-Sample and Dependent-Sample

Difference Tests," and "Three-or-More Sample Difference Tests: Analysis of Variance Methods." This reorganization highlights the contrasts that distinguish various multisample difference tests. ANOVA testing schemes receive increased emphasis.

The overall organization of the text remains unchanged from the first edition. Part one (chapters 1 and 2) introduces basic statistical concepts and terminology. Part two (chapters 3 and 4) focuses on descriptive statistics (both nonspatial and spatial) and their use in solving geographic problems. The chapters in part three (5, 6, and 7) make the transition from descriptive analysis to inferential problem solving. Topics covered include probability, sampling, and estimation. Chapters 8 through 12 make up part four of the text and cover a variety of inferential statistics for geographic problem solving. Part five (chapters 13 and 14) examines the related techniques of correlation and regression, procedures that allow measurement of association between variables and prediction of the nature of the relationship between variables. The final epilogue chapter (chapter 15) provides the list of "geographic situations" and matching statistical techniques.

Introductory quantitative methods and spatial analysis courses vary among geography programs. This textbook can be adapted to a one-semester, two-quarter, or two-semester course sequence. However the course is structured, basic instruction in the use of statistical software packages is needed.

In the context of a single semester or two quarters, several approaches to the text are possible. One alternative is to move quickly through the entire text, emphasizing those topics that seem most appropriate. Another alternative is to eliminate certain topics and chapters entirely. Instructors have considerable flexibility in determining the depth of treatment for the topics in the text. The appropriate integration of the text into a specific spatial analysis course may best be determined from the previous statistical and computing background of the students.

In the context of a two-semester (or three-quarter) sequence, the entire text can be covered in considerable detail. One option is to present parts one through three in the first semester. Students would then be well grounded in basic statistical concepts, descriptive statistics, probability, and sampling. Parts four and five would then be presented in the second semester, and students could examine the variety of inferential statistics in considerable depth.

An exercise manual accompanies the textbook. This workbook contains a variety of exercises and problems keyed directly to the material presented in the text. Geographic problems and exercises are available for topics and techniques discussed in

every chapter. To address the problems, students are referred to one of several data sets included in the workbook.

The data sets in the workbook are organized to provide maximum flexibility for the instructor and efficiency for the student. Instructors often disagree on the relative merits of requiring students to take samples from a population data set. Although sampling is an integral element of statistical problem solving, it can be a tedious and time-consuming task. The workbook offers the instructor several alternatives by identifying a few "predrawn" random samples within each data set. Different students can use different samples, avoiding the situation of every student analyzing the same set of values, or all students in the class can work with the sample data. The instructor also has the option of requiring students to draw their own sample.

ACKNOWLEDGMENTS

CONTENTS

BASIC STATISTICAL CONCEPTS IN GEOGRAPHY

Introduction: The Context of Statistical Techniques

Geography is an integrative spatial science that attempts to explain and to predict the spatial distribution and variation of human activity and physical features on the earth's surface. Geographers study how and why things differ from place to place, as well as how spatial patterns change through time. Of particular interest to geographers are the relationships between human activities and the environment and the connections between people and places. People have been interested in such geographic concerns for thousands of years. Early Greek writers, such as Eratosthenes and Strabo, emphasized the earth's physical structure, its human activity patterns, and the relationships between them.

These traditions remain central to the discipline of geography today. Contemporary geography continues to be an exciting discipline that attempts to solve a variety of problems and issues from spatial and ecological perspectives. The spatial perspective focuses on patterns and processes on the earth's surface, and the ecological perspective focuses on the complex web of relationships between living and nonliving elements on the earth's surface (Geography Education Standards Project, 1994).

The geographer starts by asking *where* questions. Where are things located on the earth's surface? How are features distributed on the physical or cultural landscape? What spatial patterns are observable, and how do phenomena vary from location to location? Historically speaking, geographers have focused on trying to answer these questions. In its popular image, the discipline of geography remains focused almost exclusively on the location of places.

However, in reality professional geographers no longer limit themselves to spatial or locational description. After a spatial pattern has been described and the where questions have been answered adequately, attention shifts to *why* questions. Why does a particular spatial pattern exist? Why does a locational pattern vary in a specific observable way? What spatial or ecological processes have affected a pattern? Why are these processes operating? As such why questions are answered, or as geographers speculate about why a spatial pattern has a particular distribution, we gain a better understanding of the processes that create the pattern. Sometimes different variables are found to be related spatially, and these findings provide insights into underlying spatial processes. In other instances, geographers try to determine if phenomena differ in various locations or regions and seek to understand why such differences exist.

Geographers are increasingly concerned with the practical application of this spatial information. Geographers ask *what-to-do* questions that involve the development of spatial policies and plans. More geographers now want to be active participants in both public and private decision making. Geographers might explore such questions as: What type of policy might best achieve more equal access for urban residents to city services and facilities? and, What sort of government policy would a geographer recommend to balance protection of wetlands and economic development in a fragile environment?

Geography is now a problem-solving discipline, and geographers are concerned with applying

their spatial knowledge and understanding to the problems facing the world today. Noted geographer Risa Palm recently stated, "[G]eography involves the study of major problems facing humankind such as environmental degradation, unequal distribution of resources and international conflicts. It prepares one to be a good citizen and educated human being" (Assoc. of American Geographers, n.d.).

1.1 THE ROLE OF STATISTICS IN GEOGRAPHY

Statistics is generally defined as the collection, classification, presentation, and analysis of numerical data. Statistical techniques and procedures are applied in all fields of academic research. In fact, wherever data are collected and summarized or wherever any numerical information is analyzed or research is conducted, statistics are needed for sound analysis and interpretation of results.

Beyond the widespread application of statistics in academic research, use of statistics can also be seen in many aspects of everyday life. Just consider for a moment this sampling of common applications of statistics:

1. In sports, statistics are frequently cited when measuring both individual and team performance. The success of a team (and the job security of the coach!) is often gauged by the team's winning percentage. Baseball fans are well-versed in such individual player statistics as batting average, earned run average, and slugging percentage. Goalies in hockey may be benched if their "goals against average" gets too high. Quarterbacks in professional football are evaluated by their quarterback rating—a multifaceted descriptive statistic with a number of components that few fans can identify.

2. Statistical procedures are used routinely in political polling and opinion gathering. When statistical sampling is applied properly, certain characteristics about a statistical population can be inferred based solely on information obtained from the sample. Political opinion polling is commonplace before any major election, and exit polling of a sample of voters is used to predict the election results as quickly as possible.

3. Statistics are widely used in business market analysis. Businesses constantly scrutinize all aspects of consumer behavior and details of consumer purchasing patterns because a fuller understanding of consumer actions and spending often translates into greater profit. Major business decisions, such as where to advertise and which markets are best to introduce a new product, depend on sample statistics. Many of us are familiar with the ratings systems used to estimate the relative popularity of various television programs. These business decisions, involving literally millions of dollars of television advertising and variable advertising rates, rely on well-designed statistical samples of the viewing population.

4. Most of us use a wide variety of statistical measures and procedures in making personal financial decisions. Whether purchasing a home, buying auto or life insurance, trying to develop a workable budget, or setting up an investment or savings plan, an understanding of different monetary terms and statistical measures is indispensable. Knowledgeable reading of stock tables and financial pages in a newspaper requires an understanding of various statistical measures. Two common examples are price-earnings ratio (the closing price of stock divided by the company's earnings per share for the latest 12-month period) and yield (current annual dividend rate divided by the closing price of a stock, expressed as a percentage).

5. Weather is an everyday practical concern for many people. They want to know which coat they should wear that morning and whether or not they will need an umbrella. Perhaps a winter storm system is approaching, and people want to know the likelihood of getting 2 to 4 inches of snow as the system moves through. A farmer may want to know the probability of getting measurable precipitation over the next two weeks as an aid to scheduling spring planting. Weather forecasting models based on statistics attempt to provide answers to such questions, often using probability estimates from similar historical weather situations.

Geographers use statistics in numerous ways. Statistical analysis benefits geographic investigation by helping answer the where, why, and what-to-do questions posed in the introductory discussion. Among many general applications, the use of statistics allows the geographer to

- describe and summarize spatial data,
- make generalizations concerning complex spatial patterns,
- estimate the likelihood or probability of outcomes for an event at a given location,
- use limited geographic data (sample) to make inferences about a larger set of geographic data (population),
- determine if the magnitude or frequency of some phenomenon differs from one location to another, and
- learn whether an actual spatial pattern matches some expected pattern.

The use of statistics must be placed within the context of a general research process. Most geographers now recognize the overall importance of statistics in research. Some may view the application of statistical methods as an essential element of any scientific geographic research, whereas others view statistical methods as one of many approaches that can be applied in geography. Whatever one's particular perspective, the general methodological shift, or "revolution," in the geographer's view of the world occurred in the late 1950s and early 1960s. During that time, geographers began to move from a qualitative description of the spatial distribution and variation of human and physical features to a quantitative analysis of the same features. As early as 1963, one observer concluded that the quantitative revolution itself was over (Burton, 1963). The application of quantitative methods (including a variety of statistical techniques) across all areas of geographic inquiry became generally accepted by geographers in the 1960s and has continued to serve as a fundamental methodological or procedural approach to much geographic research since that time.

The geographic research process and the roles of statistics in that process are summarized in a general organizational framework (figure 1.1). The left column of the figure lists the series of steps that lead to the formulation of hypotheses. These activities are done early in scientific inquiry. The right column of the figure diagrams the steps involved in scientific research after a hypothesis has been stated. Statistical procedures are involved in geographic research both before and after hypotheses are generated.

The sequence of tasks outlined in figure 1.1 is typical of the process many geographers follow when conducting research. This organizational framework is not the only mode of geographic research, however. Moreover, this framework should not be viewed as a rigid series of steps, but rather as a general, flexible guide used in geographic research.

The research process begins when a geographer identifies a worthwhile geographic problem to investigate. To recognize a productive research problem, the geographer must have background knowledge and experience in the area being studied. There is simply no substitute for having a strong background in the appropriate branch of the discipline.

Formulating a hypothesis is at the center of the research process (figure 1.1). A **hypothesis** is an unproven or unsubstantiated general statement concerning the problem under investigation. The investigator may have sufficient background knowledge or information from previous research (such as a review of the literature) to allow hypotheses to be readily developed. Perhaps a hypothesis can be formulated using a **model,** which is a simplified replication of the real world. A well-known model in geography is the spatial interaction model that predicts the amount of movement expected between two places as a function of their populations and the distance that separates them.

In some cases, a geographer may have identified a possible research area, but is not yet ready to formulate a hypothesis. Perhaps more information is needed, or questions need to be answered about the problem. In these situations, the geographer cannot immediately move down the right column of figure 1.1, but must first address research tasks in the left column. Statistical analysis is central to this process. Geographers often gain spatial insights by collecting data and presenting this information using graphical procedures and maps. During this phase of the investigation, statistical analysis also provides quantitative summaries or numerical descriptions of the data. The information gathered may enable the geographer to draw conclusions about the research questions, develop a model of the spatial situation, and generate suitable hypotheses.

If the geographer has a workable hypothesis, the research process then follows the steps shown in the right column of figure 1.1. Additional data need to be collected and prepared so that the hypothesis can be tested and evaluated. The steps following hypothesis formulation are the core of statistical analysis and are a primary focus of this book. In scientific research, if a hypothesis is repeatedly verified as correct under a variety of circumstances (perhaps at various locations or times), it gradually takes on the stature of a **law.** In other words, a proven hypothesis can eventually become accepted as a law. If various laws are combined, they then constitute a **theory.**

After testing a research hypothesis, the results must be evaluated and conclusions drawn. Several strategies or actions are possible (figure 1.1):

1. The research findings may be incorporated into actual or recommended spatial policies and plans. The applied geographer might be suggesting actions or addressing what-to-do questions.
2. If a hypothesis is verified as correct and valid, the results could be refined further into a spatial model that predicts what is likely to occur under various scenarios. Repeated verification of a hypothesis or model under a variety of circumstances (perhaps at various locations or at different times) might lead to the eventual development of laws and theories concerning the particular geographic problem or issue.
3. If a hypothesis is tested and found to be partially or completely incorrect, the geographer may need to return to an earlier step in the geographic research process. With only a partially validated
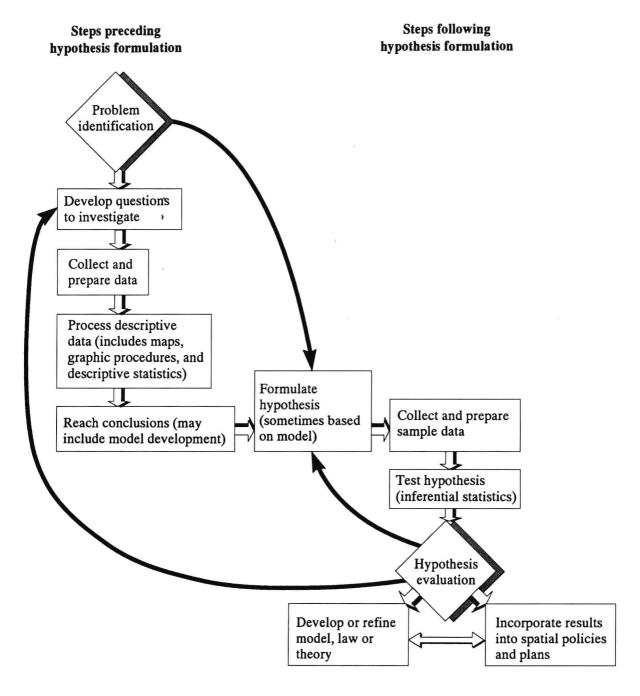
FIGURE 1.1
The Role of Statistics in the Geographic Research Process

hypothesis, one could return to hypothesis formulation to restate or refine the original hypothesis. If a hypothesis is proven totally wrong, it could be necessary to return to question development, the initial step.

The general research process outlined in figure 1.1 takes on many operational forms in geography. However, geographic research centers on the investigation of spatial geographic patterns and processes through descriptive analysis, hypothesis generation, and inferential statistical tests, with a possible goal of developing laws and theories. Collection, presentation, and processing of data all play central roles in this research process.

1.2 EXAMPLES OF STATISTICAL PROBLEM SOLVING IN GEOGRAPHY

How can geographers use statistics in the research process to approach locational issues and solve spatial problems? This will depend partly on the nature of the issue or problem under consideration and partly on the purposes and objectives of the research itself. To illustrate the diversity of approaches, two different spatial patterns are examined:

- Population change in the United States during the 1980s
- Acid deposition across North America

Both of these examples show interesting locational patterns, which are the result of complex spatial processes that are not completely understood. Figures show *where* population growth occurs in the United States and *where* acid deposition occurs in North America, but they do not explain *why* these particular spatial patterns exist.

These two examples emphasize different research needs. The first example focuses on the state-level growth rate pattern from 1980 to 1990, and discussion is closely connected to the steps in the research process (figure 1.1). The goal is to formulate a set of hypotheses based on survey responses from a typical or representative sample of interstate migrants. The second example uses an isoline map of acid deposition in North America to initiate the research process, and discussion is not tied as closely to the research process. Instead, the discussion leads to formulation of hypotheses about acid deposition patterning around a single, isolated emission source.

Population Change in the United States (1980–1990)

Suppose a geographer is interested in analyzing growth trends in the United States during the 1980s

(figure 1.2). The fastest growing states are generally located in the South and West, with much of the national growth occurring in California, Florida, and Texas. Nearly all of the states losing population or experiencing slow growth are located in the Northeast and Midwest.

Given this map of state-level population growth, the geographer may want to explore further and ask why this spatial pattern exists. Why does the growth rate pattern vary in this way? What factors can be suggested to help explain the nature of this spatial distribution? What spatial process or processes might have been operating?

Various factors may be at work to produce the spatial pattern shown in figure 1.2. Examination of previous studies may help the geographer identify potential relationships or explanations that might be relevant. For example, many geographers feel that climatic amenities and environmental considerations influence the spatial pattern of population change in the United States. The southern and western growth regions, including Florida, California, Arizona, and Hawaii, are known as the Sunbelt because of their warm and sunny climates. Conversely, the northeastern and midwestern states that are either losing population or experiencing slow growth are in regions called the Snowbelt or Frostbelt. The implication is that people are fleeing the cold winters and snow for year-round warm weather and outdoor recreation opportunities.

Economic factors undoubtedly influence growth in a number of ways. In surveys examining why people change residence, respondents frequently cite job opportunities and related economic reasons. Over the last few decades, many more new jobs have been created in the southern and western states than in the Northeast and Midwest. The region containing the states of Pennsylvania, Ohio, Michigan, Indiana, and Illinois is sometimes called the Rust Belt because of its traditional economic reliance on heavy industry and manufacturing—sectors of the U.S. economy that suffered a relative decline in the 1980s.

Patterns of immigration also influence population growth. A large number of migrants from Asia and Latin America have settled in California, Florida, and Texas, contributing to the high growth rates of these states in the 1980s. However, many immigrants have also settled in Illinois and New York, states that have experienced little recent change in population size.

Geographers studying recent migration trends suggest that low-density residential areas, such as rural regions and small towns, are increasingly attractive locations for many Americans. If this trend toward nonmetropolitan growth or decentralization