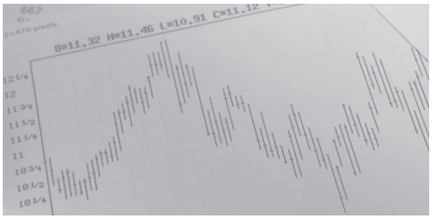

Preface

This book is written for those who need to use statistical methods to analyze data from experiments and observational studies and who need to communicate the results to others. It is intended as a text for graduate students who are preparing to design, implement, analyze, and report their research. The students must have some knowledge of basic statistical concepts such as means, standard deviations, histograms, the normal and t-distributions, but they need not be familiar with calculus or matrix algebra. All should have access to a statistical software package and a moderately powerful computer.

TO THE STUDENT

Statistics is like grout—the word feels decidedly unpleasant in the mouth, but it describes something essential for holding a mosaic in place. Statistics is a common bond supporting all other sciences. It provides standards of empirical proof and a language for communicating scientific results. Statistical sleuthing is the process of using statistical tools to answer questions of interest. It includes devising experiments to unearth hidden truths, describing real data using tools based on ideal mathematical models, answering the questions of interest efficiently, verifying that the tools are appropriate, and snooping around to see if there is anything more to be learned. *The Statistical Sleuth* will show you how this all comes about.

Case Studies

The Statistical Sleuth is organized around case studies, which begin each chapter and are used throughout to illustrate how the statistical tools operate. A small section entitled *Statistical Conclusions* accompanies each case study, demonstrating how to communicate statistical findings for a research publication. You should realize that the methods upon which the findings are based will be foreign to you upon first reading. After the chapter has been read, you should return to the studies and consider carefully how the chapter's methods have been used to answer the questions posed by the researchers.

Examine each case study carefully for its structural design. Ask yourself why the study was structured in the way it was. The studies will not only illustrate analytical techniques; most also present exemplary structures for your own studies.

Mathematical Level

The emphasis of this book is on the practical use of statistical methods. The correct practical use of statistical tools requires some understanding of the mathematical foundation for the tools. Sometimes algebra or elementary mathematical statistics are the best device for communicating the motivation. In general, however, the level of mathematics required to follow this book is not high.

What will you learn?

Do not expect to learn all that you will need to make you an experienced analyst. You will improve your understanding of statistical reasoning and of measures of uncertainty. You will learn how to translate mountains of computer output into short summary statements that communicate the results in a language common to all scientists. You will also learn a fairly large array of statistical tools that will be useful for a wide range of problems. But there are many more tools that are not covered in this book and many lessons that can only be gained through experience. At some point you may need to seek the help of a professional statistician. Then, at least, you will know the language, the general tools, and the spirit of statistical data analysis, which will make communication with a statistician more effective and beneficial.

Resources

Visit the Sleuth Web page or access course materials and companion resources at www.cengagebrain.com. At the cengagebrain.com home page, search for the ISBN of your title (from the back of your book) using the search box at the top of the page. This will take you to the product page where free companion resources can be found. Or go to www.StatisticalSleuth.com for answers to selected exercises, updates, and links to other sites: http://www.statisticalsleuth.com.

TO THE INSTRUCTOR

Level of Sophistication

The level of sophistication for this text is high when it comes to models and methods needed to analyze data and interpret results, but low when it comes to mathematics. Our foremost concern is that future researchers learn proper approaches for conducting the statistical aspects of their research. To this end, mathematics is neither sought out nor avoided.

Case Studies

Most chapters begin with two case studies for motivation and demonstration. Making these studies a central feature forces us to consider applied statistics more seriously than if we simply provided a data set to demonstrate a particular tool. It compels us to maintain a question-driven approach to the analysis of data.

The case studies are also our tool for exciting students. We cannot successfully teach them if they are not genuinely interested. We have tried to find a variety of interesting real data sets where the statistical analysis provides useful answers to

questions of interest. In some cases, we found descriptions and summary statistics that made excellent examples, but we were unable to obtain the raw data. We have still used some of them, however, by generating data to match the summary statistics. To identify these cases, we use the phrase "based on a real study."

Although we have made the data problems central, limitations of space prevent us from including all the graphical displays and the different analyses we would like to present. We encourage the instructor to go into more depth in showing computer output, graphical displays, and alternative analyses.

The Starting Point

At first glance, the first four chapters of *The Statistical Sleuth* appear to rehash the topics of one-sample and two-sample analysis, which are covered in introductory courses. This is not the case. These chapters are intended as a model for how topics in the rest of the book are treated. The chapters provide a detailed examination of material to which students have already been exposed, and an introduction to a philosophy of learning and using statistics.

Material Covered

The Statistical Sleuth's principal tool is regression analysis. We have added several topics that are not ordinarily covered in a regression text: (1) *Generalized linear models*, including logistic and log-linear regression. This important tool enables researchers to analyze a wide range of problems that have until recently been analyzed with inappropriate tools (ANOVA) or with appropriate but difficult tools (contingency table chi-squares). We stress the parallels between generalized linear model regression and ordinary regression. With calculations provided by statistical software, this tool has become entirely understandable and extremely valuable. (2) *Repeated measures*. Whereas there is a great tendency for researchers to turn to a statistical computer packaged function that has "repeated measures" in the title, we feel there needs to be more guidance on a strategy for considering such data analysis. Chapters 16 and 17 respectively emphasize question-driven and data-driven reduction of dimensionality. (3) *Serial correlation*. Although a full treatment of time series analysis is beyond the scope of this book, by adjusting and filtering for the first-order autoregression we provide tools that expand the usefulness of regression technology to problems involving serial correlation.

Our decisions regarding coverage reflect the kinds of problems graduate researchers typically encounter. The topics chosen arise repeatedly in the campus-wide consulting service operated by Oregon State University's Department of Statistics for faculty and graduate students. By offering a textbook with these topics, we hope to relieve the pressure on departments to offer separate courses in categorical data analysis, in multivariate analysis, and in time series analysis for nonstatistics majors, or to enroll nonstatistics majors in classes designed primarily for statistics majors.

Possible Paths Through the Chapters

The Statistical Sleuth was designed for a three-quarter sequence covering eight chapters each term. Typically not all the material is covered, and we have provided

a number of optional topics in a section titled *Related Issues* at the end of most chapters. The book may also be used for a two-semester class in its entirety. For a one-semester or two-quarter class, we recommend the following sequence: conclusions and interpretations (1–4), several sample problems (5–6), simple linear regression (7–8), multiple regression (9–12), two-way analysis of variance (13–14), and logistic regression (20–21). There is room to mix and match to meet specific needs.

The Statistical Sleuth covers regression prior to two-way analysis of variance, in contrast with the more traditional presentation of two-way ANOVA directly after one-way ANOVA. Our reasoning here is that regression tools applied to the two-way situation are easier to interpret. They are also less subject to misunderstandings arising from imbalance in the experimental or sampling design. Experimental design chapters (23–24) appear at the end of the book. In truth, design issues are discussed throughout the text as they apply to the case studies. Topics such as replication, blocking, factorial treatment structures, and randomization appear repeatedly. So the actual chapters on design organize and summarize the issues. We also believe it is difficult to design an experiment without an understanding of the applicable analytic tools.

Web Site

To access additional course materials and companion resources, please visit www.cengagebrain.com. At the cengagebrain.com home page, search for the ISBN of your title (from the back cover of your book) using the search box at the top of the page. This will take you to the product page where free companion resources can be found. Or go to www.StatisticalSleuth.com.

Statistical Computer Programs

A computer and a packaged statistical computer program are essential companions for *The Statistical Sleuth*. Many good packages are available. Unfortunately, they differ considerably in their style, language, and output. Some instruction about the particular software package must accompany your instruction for using the statistical tools in this book. The student's statistical analysis, however, should be guided by good statistical strategy, and not by the package, which is just a means for accomplishing the end. The data sets presented as case studies and as exercises are available online at www.cengagebrain.com.

ACKNOWLEDGMENTS

We take pleasure in expressing our gratitude to those who have helped us with this project. We are especially indebted to those who taught from earlier versions of this book. Helen Berg (mayor of Corvallis), Ginny Lesser (Oregon State University), Loretta Thielman (University of Wisconsin—Stout), and Peter Thielman (University of Wisconsin—Stout) all cheerfully endured mistakes and made valuable suggestions for improvements. Thanks also to other colleagues who read drafts and provided suggestions, including Rick Rossi, Mike Scott, and Jeannie Sifneos.

We are very grateful for the use of the facilities at the Department of Statistics at Oregon State University, and the supportive cooperation of the chairman, Justus Seely, and the assistance of Genevieve Downing. Our views of statistics have benefited from fruitful discussions with our colleagues in the Statistics Department.

A portion of the early work on the book was conducted while Dan Schafer was on sabbatical leave at The University of Western Australia. The Mathematics Department there supplied facilities for which we are grateful. Special thanks go to Ian James, currently at Murdoch University in Perth.

A very enjoyable aspect of this project for us has been the discovery of modern scientific applications we have found while searching for case studies. We have gained inspiration from the many scientists who know good questions to ask in their research and whose scientific creativity is spurred by a genuine interest in finding the answers.

Very special thanks go to the students of Statistics 511–513 classes at Oregon State University. Their maturity and collective knowledge about scientific subjects have made the course a pleasure to teach. We have been driven by a desire to provide them with tools and strategies they may use to become the scientists that inspire us in the future.

We were guided through production by the staff at Hearthside Publishing Services, Anne Seitz and Laura Horowitz.

Finally, we wish to express our gratitude to the following reviewers: Andrew Barron, Yale University; Darl Bien, University of Denver; Phil Everson, Swarthmore University; Mauro Gasparini, Purdue University; Joseph Glaz, University of Connecticut; Mark Kaiser, Iowa State University; Frank Martin, University of Minnesota; Michael Martin, Australian National University; David Mathiason, Rochester Institute of Technology; Julia Norton, California State University—Hayward; John Rawlings, North Carolina State University; Larry Ringer, Texas A&M University; Bruce Schaalje, Brigham Young University; Joseph Schafer, Pennsylvania State University; Kirk Steinhorst, University of Idaho; Loretta Thielman, University of Wisconsin—Stout; Bruce Trumbo, California State University—Hayward; and John Wasik, North Carolina State University.

Third Edition Notes

We dropped the statistical tables at the end of the book—it was time for them to go. We've added approximately 70 new data problems for exercises, as well as new sections on the Dunnett multiple comparison procedure, reasoning fallacies associated with statistical hypothesis testing, control of false discovery rates for large families of hypothesis tests, Monte Carlo methods, negative binomial regression, and generalized estimating equations. We are grateful to the following individuals for helpful comments and suggestions used in this revision: Scott Freeman, Eugene Gallagher, Alix Gitelman, Megan Higgs, Elizabeth Housworth, Martin Jones, William Kitto, Tim Leonard, Virginia Lesser, Lisa Madsen, Xiao Li Meng, Paul Murtaugh, Roger Pinkham, Berwin Turlach, Edward Whitney, and Captain William M. Wilder.

Fred Ramsey
Daniel Schafer