Flexible Imputation Of Missing Data 1st Edition

Flexible Imputation of Missing Data, Second Edition

Missing data pose challenges to real-life data analysis. Simple ad-hoc fixes, like deletion or mean imputation, only work under highly restrictive conditions, which are often not met in practice. Multiple imputation replaces each missing value by multiple plausible values. The variability between these replacements reflects our ignorance of the true (but missing) value. Each of the completed data set is then analyzed by standard methods, and the results are pooled to obtain unbiased estimates with correct confidence intervals. Multiple imputation is a general approach that also inspires novel solutions to old problems by reformulating the task at hand as a missing-data problem. This is the second edition of a popular book on multiple imputation, focused on explaining the application of methods through detailed worked examples using the MICE package as developed by the author. This new edition incorporates the recent developments in this fast-moving field. This class-tested book avoids mathematical and technical details as much as possible: formulas are accompanied by verbal statements that explain the formula in accessible terms. The book sharpens the reader’s intuition on how to think about missing data, and provides all the tools needed to execute a well-grounded quantitative analysis in the presence of missing data.

Flexible Imputation of Missing Data

Introduction -- Multiple imputation -- Univariate missing data -- Multivariate missing data -- Analysis of imputed data -- Imputation in practice -- Multilevel multiple imputation -- Individual causal effects -- Measurement issues -- Selection issues -- Longitudinal data -- Conclusion

Multiple Imputation of Missing Data in Practice

Multiple Imputation of Missing Data in Practice: Basic Theory and Analysis Strategies provides a comprehensive introduction to the multiple imputation approach to missing data problems that are often encountered in data analysis. Over the past 40 years or so, multiple imputation has gone through rapid development in both theories and applications. It is nowadays the most versatile, popular, and effective missing-data strategy that is used by researchers and practitioners across different fields. There is a strong need to better understand and learn about multiple imputation in the research and practical community. Accessible to a broad audience, this book explains statistical concepts of missing data problems and the associated terminology. It focuses on how to address missing data problems using multiple imputation. It describes the basic theory behind multiple imputation and many commonly-used models and methods. These ideas are illustrated by examples from a wide variety of missing data problems. Real data from studies with different designs and features (e.g., cross-sectional data, longitudinal data, complex surveys, survival data, studies subject to measurement error, etc.) are used to demonstrate the methods. In order for readers not only to know how to use the methods, but understand why multiple imputation works and how to choose appropriate methods, simulation studies are used to assess the performance of the multiple imputation methods. Example datasets and sample programming code are either included in the book or available at a github site (https://github.com/he-zhang-hsu/multiple_imputation_book). Key Features Provides an overview of statistical concepts that are useful for better understanding missing data problems and multiple imputation analysis Provides a detailed discussion on multiple imputation models and methods targeted to different types of missing data problems (e.g., univariate and multivariate missing data problems, missing data in survival analysis, longitudinal data, complex surveys, etc.) Explores measurement error problems with multiple imputation Discusses analysis strategies for multiple imputation diagnostics Discusses data production issues when the goal of multiple imputation is to release datasets for public use, as done by organizations that process and manage large-scale surveys with nonresponse problems For some examples, illustrative datasets and sample programming code from popular statistical packages (e.g., SAS, R, WinBUGS) are included in the book. For others, they are available at a github site (https://github.com/he-zhang-hsu/multiple_imputation_book)

Analysis of Incomplete Multivariate Data

The last two decades have seen enormous developments in statistical methods for incomplete data. The EM algorithm and its

> extensions, multiple imputation, and Markov Chain Monte Carlo provide a set of flexible and reliable tools from inference in large classes of missing-data problems. Yet, in practical terms, those developments have had surprisingly little impact on the way most data analysts handle missing values on a routine basis. Analysis of Incomplete Multivariate Data helps bridge the gap between theory and practice, making these missing-data tools accessible to a broad audience. It presents a unified, Bayesian approach to the analysis of incomplete multivariate data, covering datasets in which the variables are continuous, categorical, or both. The focus is applied, where necessary, to help readers thoroughly understand the statistical properties of those methods, and the behavior of the accompanying algorithms. All techniques are illustrated with real data examples, with extended discussion and practical advice. All of the algorithms described in this book have been implemented by the author for general use in the statistical languages S and S Plus. The software is available free of charge on the Internet.

Flexible Imputation of Missing Data

Missing data form a problem in every scientific discipline, yet the techniques required to handle them are complicated and often lacking. One of the great ideas in statistical science—multiple imputation—fills gaps in the data with plausible values, the uncertainty of which is coded in the data itself. It also solves other problems, many of which are missing data problems in disguise. Flexible Imputation of Missing Data is supported by many examples using real data taken from the author's vast experience of collaborative research, and presents a practical guide for handling missing data under the framework of multiple imputation. Furthermore, detailed guidance of implementation in R using the author’s package MICE is included throughout the book. Assuming familiarity with basic statistical concepts and multivariate methods, Flexible Imputation of Missing Data is intended for two audiences: (Bio)statisticians, epidemiologists, and methodologists in the social and health sciences Substantive researchers who do not call themselves statisticians, but who possess the necessary skills to understand the principles and to follow the recipes This graduate-tested book avoids mathematical and technical details as much as possible: formulas are accompanied by a verbal statement that explains the formula in layperson terms. Readers less concerned with the theoretical underpinnings will be able to pick up the general idea, and technical material is available for those who desire deeper understanding. The analyses can be replicated in R using a dedicated package developed by the author.

Multiple Imputation and its Application

A practical guide to analysing partially observeddata. Collecting, analysing and drawing inferences from data iscentral to research in the medical and social sciences.Unfortunately, it is rarely possible to collect all the intendeddata. The literature on inference from the resultingincomplete data is now huge, and continues to grow both asmethods are developed for large and complex data structures, and asincreasing computer power and suitable software enable researchersto apply these methods. This book focuses on a particular statistical method foranalysing and drawing inferences from incomplete data, calledMultiple Imputation (MI). MI is attractive because it is bothpractical and widely applicable. The authors aim is to clarify theissues raised by missing data, describing the rationale for MI, therelationship between the various imputation models and associatedalgorithms and its application to increasingly complex datastructures. Multiple Imputation and its Application: Discusses the issues raised by the analysis of partiallyobserved data, and the assumptions on which analyses rest. Presents a practical guide to the issues to consider whenanalysing incomplete data from both observational studies andrandomized trials. Provides a detailed discussion of the practical use of MI withreal-world examples drawn from medical and social statistics. Explores handling non-linear relationships and interactionswith multiple imputation, survival analysis, multilevel multipleimputation, sensitivity analysis via multiple imputation, usingnon-response weights with multiple imputation and doubly robustmultiple imputation. Multiple Imputation and its Application is aimed atquantitative researchers and students in the medical and socialsciences with the aim of clarifying the issues raised by theanalysis of incomplete data data, outlining the rationale for MIand describing how to consider and address the issues that arise inits application.

Statistical Analysis with Missing Data

* Emphasizes the latest trends in the field. * Includes a new chapter on evolving methods. * Provides updated or revised material in most of the chapters.

Multiple Imputation of Missing Data Using SAS

Written for users with an intermediate background in SAS programming and statistics, this book is an excellent resource for

> anyone seeking guidance on multiple imputation. It provides both theoretical background and practical solutions for those working with incomplete data sets in an engaging example-driven format.

Missing and Modified Data in Nonparametric Estimation

This book presents a systematic and unified approach for modern nonparametric treatment of missing and modified data via examples of density and hazard rate estimation, nonparametric regression, filtering signals, and time series analysis. All basic types of missing at random and not at random, biasing, truncation, censoring, and measurement errors are discussed, and their treatment is explained. Ten chapters of the book cover basic cases of direct data, biased data, nondestructive and destructive missing, survival data modified by truncation and censoring, missing survival data, stationary and nonstationary time series and processes, and ill-posed modifications. The coverage is suitable for self-study or a one-semester course for graduate students with a prerequisite of a standard course in introductory probability. Exercises of various levels of difficulty will be helpful for the instructor and self-study. The book is primarily about practically important small samples. It explains when consistent estimation is possible, and why in some cases missing data should be ignored and why others must be considered. If missing or data modification makes consistent estimation impossible, then the author explains what type of action is needed to restore the lost information. The book contains more than a hundred figures with simulated data that explain virtually every setting, claim, and development. The companion R software package allows the reader to verify, reproduce and modify every simulation and used estimators. This makes the material fully transparent and allows one to study it interactively. Sam Efromovich is the Endowed Professor of Mathematical Sciences and the Head of the Actuarial Program at the University of Texas at Dallas. He is well known for his work on the theory and application of nonparametric curve estimation and is the author of Nonparametric Curve Estimation: Methods, Theory, and Applications. Professor Sam Efromovich is a Fellow of the Institute of Mathematical Statistics and the American Statistical Association.

Handbook of Missing Data Methodology

Missing data affect nearly every discipline by complicating the statistical analysis of collected data. But since the 1990s, there have been important developments in the statistical methodology for handling missing data. Written by renowned statisticians

> in this area, Handbook of Missing Data Methodology presents many methodological advances and the latest applications of missing data methods in empirical research. Divided into six parts, the handbook begins by establishing notation and terminology. It reviews the general taxonomy of missing data mechanisms and their implications for analysis and offers a historical perspective on early methods for handling missing data. The following three parts cover various inference paradigms when data are missing, including likelihood and Bayesian methods; semi-parametric methods, with particular emphasis on inverse probability weighting; and multiple imputation methods. The next part of the book focuses on a range of approaches that assess the sensitivity of inferences to alternative, routinely non-verifiable assumptions about the missing data process. The final part discusses special topics, such as missing data in clinical trials and sample surveys as well as approaches to model diagnostics in the missing data setting. In each part, an introduction provides useful background material and an overview to set the stage for subsequent chapters. Covering both established and emerging methodologies for missing data, this book sets the scene for future research. It provides the framework for readers to delve into research and practical applications of missing data methods.

Multiple Imputation for Nonresponse in Surveys

Demonstrates how nonresponse in sample surveys and censuses can be handled by replacing each missing value with two or more multiple imputations. Clearly illustrates the advantages of modern computing to such handle surveys, and demonstrates the benefit of this statistical technique for researchers who must analyze them. Also presents the background for Bayesian and frequentist theory. After establishing that only standard complete-data methods are needed to analyze a multiply-imputed set, the text evaluates procedures in general circumstances, outlining specific procedures for creating imputations in both the ignorable and nonignorable cases. Examples and exercises reinforce ideas, and the interplay of Bayesian and frequentist ideas presents a unified picture of modern statistics.

Principles and Practice of Clinical Trials

This is a comprehensive major reference work for our SpringerReference program covering clinical trials. Although the core of the Work will focus on the design, analysis, and interpretation of scientific data from clinical trials, a broad spectrum of

> clinical trial application areas will be covered in detail. This is an important time to develop such a Work, as drug safety and efficacy emphasizes the Clinical Trials process. Because of an immense and growing international disease burden, pharmaceutical and biotechnology companies continue to develop new drugs. Clinical trials have also become extremely globalized in the past 15 years, with over 225,000 international trials ongoing at this point in time. Principles in Practice of Clinical Trials is truly an interdisciplinary that will be divided into the following areas: 1) Clinical Trials Basic Perspectives 2) Regulation and Oversight 3) Basic Trial Designs 4) Advanced Trial Designs 5) Analysis 6) Trial Publication 7) Topics Related Specific Populations and Legal Aspects of Clinical Trials The Work is designed to be comprised of 175 chapters and approximately 2500 pages. The Work will be oriented like many of our SpringerReference Handbooks, presenting detailed and comprehensive expository chapters on broad subjects. The Editors are major figures in the field of clinical trials, and both have written textbooks on the topic. There will also be a slate of 7-8 renowned associate editors that will edit individual sections of the Reference.

Clinical Trials with Missing Data

This book provides practical guidance for statisticians, clinicians, and researchers involved in clinical trials in the biopharmaceutical industry, medical and public health organisations. Academics and students needing an introduction to handling missing data will also find this book invaluable. The authors describe how missing data can affect the outcome and credibility of a clinical trial, show by examples how a clinical team can work to prevent missing data, and present the reader with approaches to address missing data effectively. The book is illustrated throughout with realistic case studies and worked examples, and presents clear and concise guidelines to enable good planning for missing data. The authors show how to handle missing data in a way that is transparent and easy to understand for clinicians, regulators and patients. New developments are presented to improve the choice and implementation of primary and sensitivity analyses for missing data. Many SAS code examples are included – the reader is given a toolbox for implementing analyses under a variety of assumptions.

The R Book

The high-level language of R is recognized as one of the mostpowerful and flexible statistical software environments, and israpidly becoming the standard setting for quantitative analysis,statistics and graphics. R provides free access to unrivalledcoverage and cutting-edge applications, enabling the user to applynumerous statistical methods ranging from simple regression to timeseries or multivariate analysis. Building on the success of the author's bestsellingStatistics: An Introduction using R, The R Book ispacked with worked examples, providing an all inclusive guide to R,ideal for novice and more accomplished users alike. The bookassumes no background in statistics or computing and introduces theadvantages of the R environment, detailing its applications in awide range of disciplines. Provides the first comprehensive reference manual for the Rlanguage, including practical guidance and full coverage of thegraphics facilities. Introduces all the statistical models covered by R, beginningwith simple classical tests such as chi-square and t-test. Proceeds to examine more advance methods, from regression andanalysis of variance, through to generalized linear models,generalized mixed models, time series, spatial statistics,multivariate statistics and much more. The R Book is aimed at undergraduates, postgraduates andprofessionals in science, engineering and medicine. It is alsoideal for students and professionals in statistics, economics,geography and the social sciences.

Multiple Imputation in Practice

Multiple Imputation in Practice: With Examples Using IVEware provides practical guidance on multiple imputation analysis, from simple to complex problems using real and simulated data sets. Data sets from cross-sectional, retrospective, prospective and longitudinal studies, randomized clinical trials, complex sample surveys are used to illustrate both simple, and complex analyses. Version 0.3 of IVEware, the software developed by the University of Michigan, is used to illustrate analyses. IVEware can multiply impute missing values, analyze multiply imputed data sets, incorporate complex sample design features, and be used for other statistical analyses framed as missing data problems. IVEware can be used under Windows, Linux, and Mac, and with software packages like SAS, SPSS, Stata, and R, or as a stand-alone tool. This book will be helpful to researchers looking for guidance on the use of multiple imputation to address missing data problems, along with examples of correct analysis techniques.

Missing Data

Missing data have long plagued those conducting applied research in the social, behavioral, and health sciences. Good missing data analysis solutions are available, but practical information about implementation of these solutions has been lacking. The objective of Missing Data: Analysis and Design is to enable investigators who are non-statisticians to implement modern missing data procedures properly in their research, and reap the benefits in terms of improved accuracy and statistical power. Missing Data: Analysis and Design contains essential information for both beginners and advanced readers. For researchers with limited missing data analysis experience, this book offers an easy-to-read introduction to the theoretical underpinnings of analysis of missing data; provides clear, step-by-step instructions for performing state-of-the-art multiple imputation analyses; and offers practical advice, based on over 20 years' experience, for avoiding and troubleshooting problems. For more advanced readers, unique discussions of attrition, non-Monte-Carlo techniques for simulations involving missing data, evaluation of the benefits of auxiliary variables, and highly cost-effective planned missing data designs are provided. The author lays out missing data theory in a plain English style that is accessible and precise. Most analysis described in the book are conducted using the well-known statistical software packages SAS and SPSS, supplemented by Norm 2.03 and associated Java-based automation utilities. A related web site contains free downloads of the supplementary software, as well as sample empirical data sets and a variety of practical exercises described in the book to enhance and reinforce the reader’s learning experience. Missing Data: Analysis and Design and its web site work together to enable beginners to gain confidence in their ability to conduct missing data analysis, and more advanced readers to expand their skill set.

The Prevention and Treatment of Missing Data in Clinical Trials

Randomized clinical trials are the primary tool for evaluating new medical interventions. Randomization provides for a fair comparison between treatment and control groups, balancing out, on average, distributions of known and unknown factors among the participants. Unfortunately, these studies often lack a substantial percentage of data. This missing data reduces the benefit provided by the randomization and introduces potential biases in the comparison of the treatment groups. Missing data can arise for a variety of reasons, including the inability or unwillingness of participants to meet appointments for evaluation. And in some studies, some or all of data collection ceases when participants discontinue study treatment. Existing

> guidelines for the design and conduct of clinical trials, and the analysis of the resulting data, provide only limited advice on how to handle missing data. Thus, approaches to the analysis of data with an appreciable amount of missing values tend to be ad hoc and variable. The Prevention and Treatment of Missing Data in Clinical Trials concludes that a more principled approach to design and analysis in the presence of missing data is both needed and possible. Such an approach needs to focus on two critical elements: (1) careful design and conduct to limit the amount and impact of missing data and (2) analysis that makes full use of information on all randomized participants and is based on careful attention to the assumptions about the nature of the missing data underlying estimates of treatment effects. In addition to the highest priority recommendations, the book offers more detailed recommendations on the conduct of clinical trials and techniques for analysis of trial data.

Handbook of International Large-Scale Assessment

Technological and statistical advances, along with a strong interest in gathering more information about the state of our educational systems, have made it possible to assess more students, in more countries, more often, and in more subject domains. The Handbook of International Large-Scale Assessment: Background, Technical Issues, and Methods of Data Analysis brings together recognized scholars in the field of ILSA, behavioral statistics, and policy to develop a detailed guide that goes beyond database user manuals. After highlighting the importance of ILSA data to policy and research, the book reviews methodological aspects and features of the studies based on operational considerations, analytics, and reporting. The book then describes methods of interest to advanced graduate students, researchers, and policy analysts who have a good grounding in quantitative methods, but who are not necessarily quantitative methodologists. In addition, it provides a detailed exposition of the technical details behind these assessments, including the test design, the sampling framework, and estimation methods, with a focus on how these issues impact analysis choices.

Statistical Rethinking

Statistical Rethinking: A Bayesian Course with Examples in R and Stan builds readers’ knowledge of and confidence in statistical modeling. Reflecting the need for even minor programming in today’s model-based statistics, the book pushes readers to perform step-by-step calculations that are usually automated. This unique computational approach ensures that

> readers understand enough of the details to make reasonable choices and interpretations in their own modeling work. The text presents generalized linear multilevel models from a Bayesian perspective, relying on a simple logical interpretation of Bayesian probability and maximum entropy. It covers from the basics of regression to multilevel models. The author also discusses measurement error, missing data, and Gaussian process models for spatial and network autocorrelation. By using complete R code examples throughout, this book provides a practical foundation for performing statistical inference. Designed for both PhD students and seasoned professionals in the natural and social sciences, it prepares them for more advanced or specialized statistical modeling. Web Resource The book is accompanied by an R package (rethinking) that is available on the author's website and GitHub. The two core functions (map and map2stan) of this package allow a variety of statistical models to be constructed from standard model formulas.

Applied Missing Data Analysis

This book has been replaced by Applied Missing Data Analysis, Second Edition, ISBN 978-1-4625-4986-3.

Multiple Imputation of Missing Data in Practice

Multiple Imputation of Missing Data in Practice: Basic Theory and Analysis Strategies provides a comprehensive introduction to the multiple imputation approach to missing data problems that are often encountered in data analysis. Over the past 40 years or so, multiple imputation has gone through rapid development in both theories and applications. It is nowadays the most versatile, popular, and effective missing-data strategy that is used by researchers and practitioners across different fields. There is a strong need to better understand and learn about multiple imputation in the research and practical community. Accessible to a broad audience, this book explains statistical concepts of missing data problems and the associated terminology. It focuses on how to address missing data problems using multiple imputation. It describes the basic theory behind multiple imputation and many commonly-used models and methods. These ideas are illustrated by examples from a wide variety of missing data problems. Real data from studies with different designs and features (e.g., cross-sectional data, longitudinal data, complex surveys, survival data, studies subject to measurement error, etc.) are used to demonstrate the methods. In order for readers not only to know how to use the methods, but understand why multiple imputation works

> and how to choose appropriate methods, simulation studies are used to assess the performance of the multiple imputation methods. Example datasets and sample programming code are either included in the book or available at a github site (https://github.com/he-zhang-hsu/multiple_imputation_book). Key Features Provides an overview of statistical concepts that are useful for better understanding missing data problems and multiple imputation analysis Provides a detailed discussion on multiple imputation models and methods targeted to different types of missing data problems (e.g., univariate and multivariate missing data problems, missing data in survival analysis, longitudinal data, complex surveys, etc.) Explores measurement error problems with multiple imputation Discusses analysis strategies for multiple imputation diagnostics Discusses data production issues when the goal of multiple imputation is to release datasets for public use, as done by organizations that process and manage large-scale surveys with nonresponse problems For some examples, illustrative datasets and sample programming code from popular statistical packages (e.g., SAS, R, WinBUGS) are included in the book. For others, they are available at a github site (https://github.com/he-zhang-hsu/multiple_imputation_book)

Bayesian Data Analysis, Third Edition

Now in its third edition, this classic book is widely considered the leading text on Bayesian methods, lauded for its accessible, practical approach to analyzing data and solving research problems. Bayesian Data Analysis, Third Edition continues to take an applied approach to analysis using up-to-date Bayesian methods. The authors—all leaders in the statistics community—introduce basic concepts from a data-analytic perspective before presenting advanced methods. Throughout the text, numerous worked examples drawn from real applications and research emphasize the use of Bayesian inference in practice. New to the Third Edition Four new chapters on nonparametric modeling Coverage of weakly informative priors and boundary-avoiding priors Updated discussion of cross-validation and predictive information criteria Improved convergence monitoring and effective sample size calculations for iterative simulation Presentations of Hamiltonian Monte Carlo, variational Bayes, and expectation propagation New and revised software code The book can be used in three different ways. For undergraduate students, it introduces Bayesian inference starting from first principles. For graduate students, the text presents effective current approaches to Bayesian modeling and computation in statistics and related fields. For researchers, it provides an assortment of Bayesian methods in applied statistics. Additional materials, including data sets used in the examples, solutions to selected exercises, and software instructions, are available on the book’s web page.

Clinical Applications of Artificial Intelligence in Real-World Data

This book is a thorough and comprehensive guide to the use of modern data science within health care. Critical to this is the use of big data and its analytical potential to obtain clinical insight into issues that would otherwise have been missed and is central to the application of artificial intelligence. It therefore has numerous uses from diagnosis to treatment. Clinical Applications of Artificial Intelligence in Real-World Data is a critical resource for anyone interested in the use and application of data science within medicine, whether that be researchers in medical data science or clinicians looking for insight into the use of these techniques.

Data Mining with SPSS Modeler

Now in its second edition, this textbook introduces readers to the IBM SPSS Modeler and guides them through data mining processes and relevant statistical methods. Focusing on step-by-step tutorials and well-documented examples that help demystify complex mathematical algorithms and computer programs, it also features a variety of exercises and solutions, as well as an accompanying website with data sets and SPSS Modeler streams. While intended for students, the simplicity of the Modeler makes the book useful for anyone wishing to learn about basic and more advanced data mining, and put this knowledge into practice. This revised and updated second edition includes a new chapter on imbalanced data and resampling techniques as well as an extensive case study on the cross-industry standard process for data mining.

Modern Analysis of Customer Surveys

Customer survey studies deals with customers, consumers and user satisfaction from a product or service. In practice, many of the customer surveys conducted by business and industry are analyzed in a very simple way, without using models or statistical methods. Typical reports include descriptive statistics and basic graphical displays. As demonstrated in this book, integrating such basic analysis with more advanced tools, provides insights on non-obvious patterns and important relationships between the survey variables. This knowledge can significantly affect the conclusions derived from a survey. Key features: Provides an integrated, case-studies based approach to analysing customer survey data. Presents a general

> introduction to customer surveys, within an organization’s business cycle. Contains classical techniques with modern and non standard tools. Focuses on probabilistic techniques from the area of statistics/data analysis and covers all major recent developments. Accompanied by a supporting website containing datasets and R scripts. Customer survey specialists, quality managers and market researchers will benefit from this book as well as specialists in marketing, data mining and business intelligence fields.

Handbook of Statistical Methods for Case-Control Studies

Handbook of Statistical Methods for Case-Control Studies is written by leading researchers in the field. It provides an in-depth treatment of up-to-date and currently developing statistical methods for the design and analysis of case-control studies, as well as a review of classical principles and methods. The handbook is designed to serve as a reference text for biostatisticians and quantitatively-oriented epidemiologists who are working on the design and analysis of case-control studies or on related statistical methods research. Though not specifically intended as a textbook, it may also be used as a backup reference text for graduate level courses. Book Sections Classical designs and causal inference, measurement error, power, and small-sample inference Designs that use full-cohort information Time-to-event data Genetic epidemiology About the Editors Ørnulf Borgan is Professor of Statistics, University of Oslo. His book with Andersen, Gill and Keiding on counting processes in survival analysis is a world classic. Norman E. Breslow was, at the time of his death, Professor Emeritus in Biostatistics, University of Washington. For decades, his book with Nick Day has been the authoritative text on case-control methodology. Nilanjan Chatterjee is Bloomberg Distinguished Professor, Johns Hopkins University. He leads a broad research program in statistical methods for modern large scale biomedical studies. Mitchell H. Gail is a Senior Investigator at the National Cancer Institute. His research includes modeling absolute risk of disease, intervention trials, and statistical methods for epidemiology. Alastair Scott was, at the time of his death, Professor Emeritus of Statistics, University of Auckland. He was a major contributor to using survey sampling methods for analyzing case-control data. Chris J. Wild is Professor of Statistics, University of Auckland. His research includes nonlinear regression and methods for fitting models to response-selective data.

Missing Data

Using numerous examples and practical tips, this book offers a nontechnical explanation of the standard methods for missing data (such as listwise or casewise deletion) as well as two newer (and, better) methods, maximum likelihood and multiple imputation. Anyone who has relied on ad-hoc methods that are statistically inefficient or biased will find this book a welcome and accessible solution to their problems with handling missing data.

Statistical Methods for Handling Incomplete Data

Due to recent theoretical findings and advances in statistical computing, there has been a rapid development of techniques and applications in the area of missing data analysis. Statistical Methods for Handling Incomplete Data covers the most up-to-date statistical theories and computational methods for analyzing incomplete data. Features Uses the mean score equation as a building block for developing the theory for missing data analysis Provides comprehensive coverage of computational techniques for missing data analysis Presents a rigorous treatment of imputation techniques, including multiple imputation fractional imputation Explores the most recent advances of the propensity score method and estimation techniques for nonignorable missing data Describes a survey sampling application Updated with a new chapter on Data Integration Now includes a chapter on Advanced Topics, including kernel ridge regression imputation and neural network model imputation The book is primarily aimed at researchers and graduate students from statistics, and could be used as a reference by applied researchers with a good quantitative background. It includes many real data examples and simulated examples to help readers understand the methodologies.

Computational Learning Approaches to Data Analytics in Biomedical Applications

Computational Learning Approaches to Data Analytics in Biomedical Applications provides a unified framework for biomedical data analysis using varied machine learning and statistical techniques. It presents insights on biomedical data processing, innovative clustering algorithms and techniques, and connections between statistical analysis and clustering. The book introduces and discusses the major problems relating to data analytics, provides a review of influential and state-of-the-art

> learning algorithms for biomedical applications, reviews cluster validity indices and how to select the appropriate index, and includes an overview of statistical methods that can be applied to increase confidence in the clustering framework and analysis of the results obtained. Includes an overview of data analytics in biomedical applications and current challenges Updates on the latest research in supervised learning algorithms and applications, clustering algorithms and cluster validation indices Provides complete coverage of computational and statistical analysis tools for biomedical data analysis Presents hands-on training on the use of Python libraries, MATLAB® tools, WEKA, SAP-HANA and R/Bioconductor

Multiple Imputation and its Application

Multiple Imputation and its Application The most up-to-date edition of a bestselling guide to analyzing partially observed data In this comprehensively revised Second Edition of Multiple Imputation and its Application, a team of distinguished statisticians delivers an overview of the issues raised by missing data, the rationale for multiple imputation as a solution, and the practicalities of applying it in a multitude of settings. With an accessible and carefully structured presentation aimed at quantitative researchers, Multiple Imputation and its Application is illustrated with a range of examples and offers key mathematical details. The book includes a wide range of theoretical and computer-based exercises, tested in the classroom, which are especially useful for users of R or Stata. Readers will find: A comprehensive overview of one of the most effective and popular methodologies for dealing with incomplete data sets Careful discussion of key concepts A range of examples illustrating the key ideas Practical advice on using multiple imputation Exercises and examples designed for use in the classroom and/or private study Written for applied researchers looking to use multiple imputation with confidence, and for methods researchers seeking an accessible overview of the topic, Multiple Imputation and its Application will also earn a place in the libraries of graduate students undertaking quantitative analyses.

Best Practices in Quantitative Methods

The contributors to Best Practices in Quantitative Methods envision quantitative methods in the 21st century, identify the best practices, and, where possible, demonstrate the superiority of their recommendations empirically. Editor Jason W. Osborne designed this book with the goal of providing readers with the most effective, evidence-based, modern quantitative methods

> and quantitative data analysis across the social and behavioral sciences. The text is divided into five main sections covering select best practices in Measurement, Research Design, Basics of Data Analysis, Quantitative Methods, and Advanced Quantitative Methods. Each chapter contains a current and expansive review of the literature, a case for best practices in terms of method, outcomes, inferences, etc., and broad-ranging examples along with any empirical evidence to show why certain techniques are better. Key Features: Describes important implicit knowledge to readers: The chapters in this volume explain the important details of seemingly mundane aspects of quantitative research, making them accessible to readers and demonstrating why it is important to pay attention to these details. Compares and contrasts analytic techniques: The book examines instances where there are multiple options for doing things, and make recommendations as to what is the \"best\" choice—or choices, as what is best often depends on the circumstances. Offers new procedures to update and explicate traditional techniques: The featured scholars present and explain new options for data analysis, discussing the advantages and disadvantages of the new procedures in depth, describing how to perform them, and demonstrating their use. Intended Audience: Representing the vanguard of research methods for the 21st century, this book is an invaluable resource for graduate students and researchers who want a comprehensive, authoritative resource for practical and sound advice from leading experts in quantitative methods.

Classification, Clustering, and Data Mining Applications

Modern data analysis stands at the interface of statistics, computer science, and discrete mathematics. This volume describes new methods in this area, with special emphasis on classification and cluster analysis. Those methods are applied to problems in information retrieval, phylogeny, medical diagnosis, microarrays, and other active research areas.

Handbook of Computational Social Science, Volume 1

The Handbook of Computational Social Science is a comprehensive reference source for scholars across multiple disciplines. It outlines key debates in the field, showcasing novel statistical modeling and machine learning methods, and draws from specific case studies to demonstrate the opportunities and challenges in CSS approaches. The Handbook is divided into two volumes written by outstanding, internationally renowned scholars in the field. This first volume focuses on the scope of

> computational social science, ethics, and case studies. It covers a range of key issues, including open science, formal modeling, and the social and behavioral sciences. This volume explores major debates, introduces digital trace data, reviews the changing survey landscape, and presents novel examples of computational social science research on sensing social interaction, social robots, bots, sentiment, manipulation, and extremism in social media. The volume not only makes major contributions to the consolidation of this growing research field but also encourages growth in new directions. With its broad coverage of perspectives (theoretical, methodological, computational), international scope, and interdisciplinary approach, this important resource is integral reading for advanced undergraduates, postgraduates, and researchers engaging with computational methods across the social sciences, as well as those within the scientifi c and engineering sectors.

Mathematical and Statistical Skills in the Biopharmaceutical Industry

Mathematical and Statistical Skills in the Biopharmaceutical Industry: A Pragmatic Approach describes a philosophy of efficient problem solving showcased using examples pertinent to the biostatistics function in clinical drug development. It was written to share a quintessence of the authors' experiences acquired during many years of relevant work in the biopharmaceutical industry. The book will be useful will be useful for biopharmaceutical industry statisticians at different seniority levels and for graduate students who consider a biostatistics-related career in this industry. Features: Describes a system of principles for pragmatic problem solving in clinical drug development. Discusses differences in the work of a biostatistician in small pharma and big pharma. Explains the importance/relevance of statistical programming and data management for biostatistics and necessity for integration on various levels. Describes some useful statistical background that can be capitalized upon in the drug development enterprise. Explains some hot topics and current trends in biostatistics in simple, non-technical terms. Discusses incompleteness of any system of standard operating procedures, rules and regulations. Provides a classification of scoring systems and proposes a novel approach for evaluation of the safety outcome for a completed randomized clinical trial. Presents applications of the problem solving philosophy in a highly problematic transfusion field where many investigational compounds have failed. Discusses realistic planning of open-ended projects.

Your Statistical Consultant

How do you bridge the gap between what you learned in your statistics course and the questions you want to answer in your real-world research? Oriented towards distinct questions in a \"How do I?\" or \"When should I?\" format, Your Statistical Consultant is the equivalent of the expert colleague down the hall who fields questions about describing, explaining, and making recommendations regarding thorny or confusing statistical issues. The book serves as a compendium of statistical knowledge, both theoretical and applied, that addresses the questions most frequently asked by students, researchers and instructors. Written to be responsive to a wide range of inquiries and levels of expertise, the book is flexibly organized so readers can either read it sequentially or turn directly to the sections that correspond to their concerns.

Handbook of Multilevel Analysis

This book presents the state of the art in multilevel analysis, with an emphasis on more advanced topics. These topics are discussed conceptually, analyzed mathematically, and illustrated by empirical examples. Multilevel analysis is the statistical analysis of hierarchically and non-hierarchically nested data. The simplest example is clustered data, such as a sample of students clustered within schools. Multilevel data are especially prevalent in the social and behavioral sciences and in the biomedical sciences. The chapter authors are all leading experts in the field. Given the omnipresence of multilevel data in the social, behavioral, and biomedical sciences, this book is essential for empirical researchers in these fields.

Statistical Analysis with Missing Data

An up-to-date, comprehensive treatment of a classic text on missing data in statistics The topic of missing data has gained considerable attention in recent decades. This new edition by two acknowledged experts on the subject offers an up-to-date account of practical methodology for handling missing data problems. Blending theory and application, authors Roderick Little and Donald Rubin review historical approaches to the subject and describe simple methods for multivariate analysis with missing values. They then provide a coherent theory for analysis of problems based on likelihoods derived from statistical models for the data and the missing data mechanism, and then they apply the theory to a wide range of important missing

> data problems. Statistical Analysis with Missing Data, Third Edition starts by introducing readers to the subject and approaches toward solving it. It looks at the patterns and mechanisms that create the missing data, as well as a taxonomy of missing data. It then goes on to examine missing data in experiments, before discussing complete-case and available-case analysis, including weighting methods. The new edition expands its coverage to include recent work on topics such as nonresponse in sample surveys, causal inference, diagnostic methods, and sensitivity analysis, among a host of other topics. An updated “classic” written by renowned authorities on the subject Features over 150 exercises (including many new ones) Covers recent work on important methods like multiple imputation, robust alternatives to weighting, and Bayesian methods Revises previous topics based on past student feedback and class experience Contains an updated and expanded bibliography The authors were awarded The Karl Pearson Prize in 2017 by the International Statistical Institute, for a research contribution that has had profound influence on statistical theory, methodology or applications. Their work \"has been no less than defining and transforming.\" (ISI) Statistical Analysis with Missing Data, Third Edition is an ideal textbook for upper undergraduate and/or beginning graduate level students of the subject. It is also an excellent source of information for applied statisticians and practitioners in government and industry.

Semiparametric Theory and Missing Data

This book summarizes current knowledge regarding the theory of estimation for semiparametric models with missing data, in an organized and comprehensive manner. It starts with the study of semiparametric methods when there are no missing data. The description of the theory of estimation for semiparametric models is both rigorous and intuitive, relying on geometric ideas to reinforce the intuition and understanding of the theory. These methods are then applied to problems with missing, censored, and coarsened data with the goal of deriving estimators that are as robust and efficient as possible.

Practical Multivariate Analysis

This is the sixth edition of a popular textbook on multivariate analysis. Well-regarded for its practical and accessible approach, with excellent examples and good guidance on computing, the book is particularly popular for teaching outside statistics, i.e. in epidemiology, social science, business, etc. The sixth edition has been updated with a new chapter on data

> visualization, a distinction made between exploratory and confirmatory analyses and a new section on generalized estimating equations and many new updates throughout. This new edition will enable the book to continue as one of the leading textbooks in the area, particularly for non-statisticians. Key Features: Provides a comprehensive, practical and accessible introduction to multivariate analysis. Keeps mathematical details to a minimum, so particularly geared toward a non-statistical audience. Includes lots of detailed worked examples, guidance on computing, and exercises. Updated with a new chapter on data visualization.

Multiple Imputation for Nonresponse in Surveys

Demonstrates how nonresponse in sample surveys and censuses can be handled by replacing each missing value with two or more multiple imputations. Clearly illustrates the advantages of modern computing to such handle surveys, and demonstrates the benefit of this statistical technique for researchers who must analyze them. Also presents the background for Bayesian and frequentist theory. After establishing that only standard complete-data methods are needed to analyze a multiply-imputed set, the text evaluates procedures in general circumstances, outlining specific procedures for creating imputations in both the ignorable and nonignorable cases. Examples and exercises reinforce ideas, and the interplay of Bayesian and frequentist ideas presents a unified picture of modern statistics.