Design And Analysis Of Cluster Randomization Trials In Health Research 1st Edition

This dissertation is focused on the development of the optimal design and analysis for cluster randomized trials. Specifically, we tackle three common questions: whether or not to pair-match clusters, which causal parameter best captures the intervention effect, and how to select the adjustment set for the analysis. We begin by introducing a formal framework for causal inference in Chapter 1. Throughout, the Sustainable East Africa Research in Community Health (SEARCH) trial serves as the motivating example (NCT01864603). SEARCH is an ongoing community randomized trial to evaluate the impact of immediate and streamlined antiretroviral therapy on HIV incidence in rural East Africa. In Chapter 2, we consider pair-matching, an intuitive design strategy to protect study validity and to potentially increase power in randomized trials. In a common design, candidate units are identified, and their baseline characteristics are used to create the best n/2 matched pairs. Within the resulting pairs, the intervention is randomized, and the outcomes are measured at the end of follow-up. We consider this design to be adaptive, because the construction of the matched pairs depends on the baseline

covariates of all candidate units. As a consequence, the observed data cannot be considered as n/2 independent, identically distributed (i.i.d.) pairs of units, as common practice assumes. Instead, the observed data consist of n dependent units. Chapter 2 explores the consequences of adaptive pair-matching in randomized trials for estimation of the conditional average treatment effect (CATE): the intervention effect, given the measured covariates of the n study units. We contrast the unadjusted estimator with TMLE and show substantial efficiency gains from matching and further gains with adjustment. In Chapter 3, we compare three causal parameters: the population, conditional and sample average treatment effects. Using a structural causal model, we explicitly define each parameter, discuss interpretation, and formally examine identifiability. To the best of our knowledge, Chapter 3 is the first to propose using TMLE for estimation and inference of the sample effect. In most settings, the sample parameter will be estimated more efficiently than the conditional parameter, which will, in turn, be estimated more efficiently than the population parameter. Finite sample simulations illustrate the potential gains in precision and power from selecting the sample effect as the target of inference. Finally in Chapter 4, we discuss adjustment for measured covariates during the analysis to reduce variance and increase power in randomized trials. To avoid

misleading inference, the analysis plan must be pre-specified. However, it is often unclear a priori which baseline covariates (if any) should be included in the analysis. In the SEARCH trial, for example, there are 16 matched pairs of communities and many potential adjustment variables, including region, HIV prevalence, male circumcision coverage and measures of community-level viral load. In Chapter 4, we propose a rigorous procedure to data-adaptively select the adjustment set, which maximizes the efficiency of the analysis. Specifically, we use cross-validation to select from a pre-specified library the candidate TMLE that minimizes the estimated variance. For further gains in precision, we also propose a collaborative procedure for estimating the known exposure mechanism. Our small sample simulations demonstrate the promise of the methodology to maximize study power, while maintaining nominal confidence interval coverage. Our procedure is tailored to the scientific question (sample vs. population treatment effect) and study design (pair-matched or not) and alleviates many of the common concerns.

Recently many researchers are working on cluster analysis as a main tool for exploratory data analysis and data mining. A notable feature is that specialists in di?erent ?elds of sciences are considering the tool of data clustering to be useful. A major reason is that clustering algorithms and software are ?exible in the

sensethatdi?erentmathematicalframeworksareemployedinthealgorithms and a user can select a suitable method according to his application. Moreover clustering algorithmshavedi?erentoutputsrangingfromtheold dendrogramsof agglomerativeclustering to more recent self-organizingmaps. Thus, a researcher or user can choose an appropriate output suited to his purpose,which is another ?exibility of the methods of clustering. An old and still most popular method is the K-means which use K cluster centers. A group of data is gathered around a cluster center and thus forms a cluster. The main subject of this book is the fuzzy c-means proposed by Dunn and Bezdek and their variations including recent studies. A main reasonwhy we concentrate on fuzzy c-means is that most methodology and application studies infuzzy clusteringusefuzzy c-means,andfuzzy c-meansshouldbe consideredto beamajortechniqueofclusteringingeneral,regardlesswhetheroneisinterested in fuzzy methods or not. Moreover recent advances in clustering techniques are rapid and we requirea new textbook that includes recent algorithms.We should also note that several books have recently been published but the contents do not include some methods studied herein.

Cluster Analysis for Applications deals with methods and various applications of cluster analysis. Topics covered range from variables and scales to measures of association among variables and

among data units. Conceptual problems in cluster analysis are discussed, along with hierarchical and non-hierarchical clustering methods. The necessary elements of data analysis, statistics, cluster analysis, and computer implementation are integrated vertically to cover the complete path from raw data to a finished analysis. Comprised of 10 chapters, this book begins with an introduction to the subject of cluster analysis and its uses as well as category sorting problems and the need for cluster analysis algorithms. The next three chapters give a detailed account of variables and association measures, with emphasis on strategies for dealing with problems containing variables of mixed types. Subsequent chapters focus on the central techniques of cluster analysis with particular reference to computational considerations; interpretation of clustering results; and techniques and strategies for making the most effective use of cluster analysis. The final chapter suggests an approach for the evaluation of alternative clustering methods. The presentation is capped with a complete set of implementing computer programs listed in the Appendices to make the use of cluster analysis as painless and free of mechanical error as is possible. This monograph is intended for students and workers who have encountered the notion of cluster analysis.

This book provides the reader with a basic understanding of the formal concepts of the cluster,

clustering, partition, cluster analysis etc. The book explains feature-based, graph-based and spectral clustering methods and discusses their formal similarities and differences. Understanding the related formal concepts is particularly vital in the epoch of Big Data; due to the volume and characteristics of the data, it is no longer feasible to predominantly rely on merely viewing the data when facing a clustering problem. Usually clustering involves choosing similar objects and grouping them together. To facilitate the choice of similarity measures for complex and big data, various measures of object similarity, based on quantitative (like numerical measurement results) and qualitative features (like text), as well as combinations of the two, are described, as well as graph-based similarity measures for (hyper) linked objects and measures for multilayered graphs. Numerous variants demonstrating how such similarity measures can be exploited when defining clustering cost functions are also presented. In addition, the book provides an overview of approaches to handling large collections of objects in a reasonable time. In particular, it addresses grid-based methods, sampling methods, parallelization via Map-Reduce, usage of tree-structures, random projections and various heuristic approaches, especially those used for community detection.

Design and Analysis of Cluster Randomization Trials

in Health ResearchWiley
Cluster analysis is used in data mining and is a common technique for statistical data analysis used in many fields of study, such as the medical & life sciences, behavioral & social sciences, engineering, and in computer science. Designed for training industry professionals or for a course on clustering and classification, it can also be used as a companion text for applied statistics. No previous experience in clustering or data mining is assumed. Informal algorithms for clustering data and interpreting results are emphasized. In order to evaluate the results of clustering and to explore data, graphical methods and data structures are used for representing data. Throughout the text, examples and references are provided, in order to enable the material to be comprehensible for a diverse audience. A companion disc includes numerous appendices with programs, data, charts, solutions, etc. eBook Customers: Companion files are available for downloading with order number/proof of purchase by writing to the publisher at info@merclearning.com. FEATURES *Places emphasis on illustrating the underlying logic in making decisions during the cluster analysis *Discusses the related applications of statistic, e.g., Ward's method (ANOVA), JAN (regression analysis & correlational analysis), cluster validation (hypothesis testing, goodness-of-fit, Monte Carlo

simulation, etc.) *Contains separate chapters on JAN and the clustering of categorical data *Includes a companion disc with solutions to exercises, programs, data sets, charts, etc.

A complete guide to understanding cluster randomised trials Written by two researchers with extensive experience in the field, this book presents a complete guide to the design, analysis and reporting of cluster randomised trials. It spans a wide range of applications: trials in developing countries, trials in primary care, trials in the health services. A key feature is the use of R code and code from other popular packages to plan and analyse cluster trials, using data from actual trials. The book contains clear technical descriptions of the models used, and considers in detail the ethics involved in such trials and the problems in planning them. For readers and students who do not intend to run a trial but wish to be a critical reader of the literature, there are sections on the CONSORT statement, and exercises in reading published trials. Written in a clear, accessible style Features real examples taken from the authors' extensive practitioner experience of designing and analysing clinical trials Demonstrates the use of R, Stata and SPSS for statistical analysis Includes computer code so the reader can replicate all the analyses Discusses neglected areas such as ethics and practical issues in running cluster randomised trials How to Design, Analyse and Report Cluster Randomised Trials in Medicine and Health Related Research provides an excellent reference tool and can be read with profit by statisticians, health services researchers, systematic reviewers and critical readers of cluster randomised trials.

Although there are several good books on unsupervised machine learning, we felt that many of them are too theoretical. This book provides practical guide to cluster analysis, elegant visualization and interpretation. It contains 5

parts. Part I provides a quick introduction to R and presents required R packages, as well as, data formats and dissimilarity measures for cluster analysis and visualization. Part II covers partitioning clustering methods, which subdivide the data sets into a set of k groups, where k is the number of groups pre-specified by the analyst. Partitioning clustering approaches include: K-means, K-Medoids (PAM) and CLARA algorithms. In Part III, we consider hierarchical clustering method, which is an alternative approach to partitioning clustering. The result of hierarchical clustering is a tree-based representation of the objects called dendrogram. In this part, we describe how to compute, visualize, interpret and compare dendrograms. Part IV describes clustering validation and evaluation strategies, which consists of measuring the goodness of clustering results. Among the chapters covered here, there are: Assessing clustering tendency, Determining the optimal number of clusters, Cluster validation statistics, Choosing the best clustering algorithms and Computing p-value for hierarchical clustering. Part V presents advanced clustering methods, including: Hierarchical k-means clustering, Fuzzy clustering, Model-based clustering and Density-based clustering.

Clustering techniques are increasingly being put to use in the analysis of high-throughput biological datasets. Novel computational techniques to analyse high throughput data in the form of sequences, gene and protein expressions, pathways, and images are becoming vital for understanding diseases and future drug discovery. This book details the complete pathway of cluster analysis, from the basics of molecular biology to the generation of biological knowledge. The book also presents the latest clustering methods and clustering validation, thereby offering the reader a comprehensive review of clustering analysis in bioinformatics from the fundamentals through to state-of-the-art techniques

and applications. Key Features: Offers a contemporary review of clustering methods and applications in the field of bioinformatics, with particular emphasis on gene expression analysis Provides an excellent introduction to molecular biology with computer scientists and information engineering researchers in mind, laying out the basic biological knowledge behind the application of clustering analysis techniques in bioinformatics Explains the structure and properties of many types of high-throughput datasets commonly found in biological studies Discusses how clustering methods and their possible successors would be used to enhance the pace of biological discoveries in the future Includes a companion website hosting a selected collection of codes and links to publicly available datasets

Accurate sample size calculation ensures that clinical studies have adequate power to detect clinically meaningful effects. This results in the efficient use of resources and avoids exposing a disproportionate number of patients to experimental treatments caused by an overpowered study. Sample Size Calculations for Clustered and Longitudinal Outcomes in Clinical Research explains how to determine sample size for studies with correlated outcomes, which are widely implemented in medical, epidemiological, and behavioral studies. The book focuses on issues specific to the two types of correlated outcomes: longitudinal and clustered. For clustered studies, the authors provide sample size formulas that accommodate variable cluster sizes and within-cluster correlation. For longitudinal studies, they present sample size formulas to account for within-subject correlation among repeated measurements and various missing data patterns. For multiple levels of clustering, the level at which to perform randomization actually becomes a design parameter. The authors show how this can greatly impact trial administration, analysis, and sample size

requirement. Addressing the overarching theme of sample size determination for correlated outcomes, this book provides a useful resource for biostatisticians, clinical investigators, epidemiologists, and social scientists whose research involves trials with correlated outcomes. Each chapter is self-contained so readers can explore topics relevant to their research projects without having to refer to other chapters.

This text provides the most comprehensive treatment of the design and analytic issues involved in group-randomized trials. GRTs are comparative studies conducted to evaluate the effect of a health promotion intervention in which the units of assignment are identifiable groups (e.g., schools, worksites) and the units of observation are members of those groups (e.g., students, workers). The book reviews the underlying issues, the most widely used research designs, and analytic strategies. There is an emphasis on mixed-model regression, with two chapters illustrating the analytic methods in SAS PROC MIXED and GLIMMIX. There is also a detailed chapter on power analysis and sample size calculation.

In this study, we conducted simulations to evaluate the effect of constrained randomization on testing the treatment effect in terms of type I error and power with data generated from a stepped wedge cluster-randomized design, under the presence of cluster-level covariates. We considered two cases, one with a single binary co- variate and the other with a mixture of continuous and categorical covariates. For case one we used stratified randomization to achieve perfect covariate balance whereas for case two we constrained the randomization space by setting scores based on different balance criteria. For each case we consider eight different scenarios, and apply model-based and permutation-based inference to estimate and test for the treatment effect, both

adjusted and unadjusted for covariates in the analysis phase. We found that the type I error is close to the nominal level most of the time except for permutation inference with constrained randomization and unconstrained analysis, where it drops down towards zero. In general, we see that constrained randomization can slightly increase power when covariates are also included at the analysis phase, and such increase is more visible in case one with a single binary covariate than case two with multiple covariates. Overall, although we discovered some advantages of doing constrained randomization in terms of power, such gain is only marginal and its impact in practice is likely to be much smaller than under traditional cluster-randomized design. Therefore, controlling for covariates in the analysis phase is still considered to be a more effective way to attain higher testing power under stepped wedge setting.

Clustering remains a vibrant area of research in statistics. Although there are many books on this topic, there are relatively few that are well founded in the theoretical aspects. In Robust Cluster Analysis and Variable Selection, Gunter Ritter presents an overview of the theory and applications of probabilistic clustering and variable selection, synthesizing the key research results of the last 50 years. The author focuses on the robust clustering methods he found to be the most useful on simulated data and real-time applications. The book provides clear guidance for the varying needs of both applications, describing scenarios in which accuracy and speed are the primary goals. Robust Cluster Analysis and Variable Selection includes all of the important theoretical details, and covers the key probabilistic models, robustness issues, optimization algorithms,

validation techniques, and variable selection methods. The book illustrates the different methods with simulated data and applies them to real-world data sets that can be easily downloaded from the web. This provides you with guidance in how to use clustering methods as well as applicable procedures and algorithms without having to understand their probabilistic fundamentals.

This volume provides a concise introduction to the various types of clustering methods typically used in the social sciences.

The stepped wedge (SW) design is a variation on the cluster randomized trial (CRT) in which all clusters receive both the control and intervention conditions. Clusters begin the trial at the same time, and all participants are assigned to the control condition during this initial phase in the study. At sequential time periods, individual clusters or groups of clusters cross over to the intervention condition based on a randomly assigned cluster order. The SW design is becoming increasingly popular due to its advantages over CRTs when intra-cluster correlation is high, and because it is a useful design in the field of implementation science. This research focused on three different design issues that statisticians and researchers are currently facing: (1) variability in cluster size and enrollment over time leading to imbalance in number of participants by treatment arm, and subsequent effects on statistical power, (2) covariate-constrained randomization as a way to mitigate potential covariate imbalance across treatment arm, and (3) efficient control group selection procedures from electronic medical records. Some of our research

questions were inspired by the DECIDE-LVAD trial, which employed a SW design, and summary data from this trial are used in examples throughout. SW designs can end up with an imbalance in the number of participants by treatment arm if the clusters are of variable size, or if participants do not enroll at a constant rate across time period within cluster. We explored the effects of variable cluster size, variable enrollment over time and treatment group imbalance on the power of SW designs. We found that increasing cluster size variability and enrollment variability across time period led to decreased power relative to designs with equal cluster-period sample sizes. To understand the difference in power between complete SW designs and those with a washout period prior to intervention implementation, we derived an expression for the variance of the treatment effect in the presence of a washout period. This showed that the difference in power between complete and washout designs is dependent on the number of treatment sequences in the SW design. Covariate-constrained randomization is commonly used in the CRT setting if there are concerns about potential confounding due to differing cluster level covariates or individual-level covariates where the distributions vary by cluster. We developed a method for covariate-constrained randomization in SW designs that is easy to implement, including shareable code. We evaluated this method by comparing treatment effect estimation, power, and type I error across analysis methods and using different constraint thresholds. We compared analysis methods in two ways: (a) linear mixed model analyses vs.

permutation tests, and (b) unadjusted vs. adjusted for potential confounders. We observed consistently good results when the covariate-constrained randomization procedure was used to rule out a small set of potential randomizations with the worst balance and final statistical analyses were adjusted for potential confounders. Recently, questions have been raised regarding efficient control group selection methods in a situation when participants in the intervention arm of the SW design will be enrolled as usual, but all outcome data will be collected from electronic health records. In this setting, control group participants would not need to be enrolled. This design variant is of interest to researchers due to the potential cost savings of enrolling a smaller group of participants. Research questions include (a) how to select the potential controls to match the risk profile of the intervention group participants, and (b) how to analyze the data after control participants are selected. First, we modified a propensity score (PS) matching technique that has been used in the clustered data setting for use in SW, and compared its matching properties to the greedy nearest neighbor matching algorithm. Next, we compared results of statistical analyses after selecting the control group using the matching methods to analysis methods which made use of the entire pool of potential controls (PS weighting, covariate adjustment). We observed that the analysis methods which employed covariate adjustment on the sample of intervention participants and the entire pool of potential controls were the most robust, even in the presence of cluster-level unmeasured confounding. This

research will be of use to researchers who are considering employing the SW design in a future trial. Our findings demonstrate the sensitivity of the power to cluster-period sample size variability, emphasize the importance of considering all potential confounders prior to cluster order randomization, and describe appropriate methods for design and analysis in the presence of potential confounding.

Research has generated a number of advances in methods for spatial cluster modelling in recent years, particularly in the area of Bayesian cluster modelling. Along with these advances has come an explosion of interest in the potential applications of this work, especially in epidemiology and genome research. In one integrated volume, this book reviews the state-of-the-art in spatial clustering and spatial cluster modelling, bringing together research and applications previously scattered throughout the literature. It begins with an overview of the field, then presents a series of chapters that illuminate the nature and purpose of cluster modelling within different application areas, including astrophysics, epidemiology, ecology, and imaging. The focus then shifts to methods, with discussions on point and object process modelling, perfect sampling of cluster processes, partitioning in space and space-time, spatial and spatio-temporal process modelling, nonparametric methods for clustering, and spatio-temporal cluster modelling. Many figures, some in full color, complement the text, and a single section of references cited makes it easy to locate source material. Leading specialists in the field of cluster modelling authored each chapter, and an

introduction by the editors to each chapter provides a cohesion not typically found in contributed works. Spatial Cluster Modelling thus offers a singular opportunity to explore this exciting new field, understand its techniques, and apply them in your own research.

Field experiments conducted with the village, city, state, region, or even country as the unit of randomization are becoming commonplace in the social sciences. While convenient, subsequent data analysis may be complicated by the constraint on the number of clusters in treatment and control. Through a battery of Monte Carlo simulations, we examine best practices for estimating unit-level treatment effects in cluster-randomized field experiments, particularly in settings that generate short panel data. In most settings we consider, unit-level estimation with unit fixed effects and cluster-level estimation weighted by the number of units per cluster tend to be robust to potentially problematic features in the data while giving greater statistical power. Using insights from our analysis, we evaluate the effect of a unique field experiment: a nationwide tipping field experiment across markets on the Uber app. Beyond the import of showing how tipping affects aggregate market outcomes, we provide several insights on aspects of generating and analyzing cluster-randomized experimental data when there are constraints on the number of experimental units in treatment and control.

Cluster randomised trials are trials in which groups (or clusters) of individuals are randomly allocated to different forms of treatment. In health care, these trials often compare different ways of managing a disease or

promoting healthy living, in contrast to conventional randomised trials which randomise individuals to different treatments, classically comparing new drugs with a placebo. They are increasingly common in health services research. This book addresses the statistical, practical, and ethical issues arising from allocating groups of individuals, or clusters, to different interventions. Key features: Guides readers through the stages of conducting a trial, from recruitment to reporting. Presents a wide range of examples with particular emphasis on trials in health services research and primary care, with both principles and techniques explained. Topics are specifically presented in the order in which investigators think about issues when they are designing a trial. Combines information on the latest developments in the field together with a practical guide to the design and implementation of cluster randomised trials. Explains principles and techniques through numerous examples including many from the authors own experience. Includes a wide range of references for those who wish to read further. This book is intended as a practical guide, written for researchers from the health professions including doctors, psychologists, and allied health professionals, as well as statisticians involved in the design, execution, analysis and reporting of cluster randomised trials. Those with a more general interest will find the plentiful examples illuminating.

The Wiley-Interscience Paperback Series consists of selected books that have been made more accessible to consumers in an effort to increase global appeal and general circulation. With these new unabridged softcover

volumes, Wiley hopes to extend the lives of these works by making them available to future generations of statisticians, mathematicians, and scientists. "Cluster analysis is the increasingly important and practical subject of finding groupings in data. The authors set out to write a book for the user who does not necessarily have an extensive background in mathematics. They succeed very well." —Mathematical Reviews "Finding Groups in Data [is] a clear, readable, and interesting presentation of a small number of clustering methods. In addition, the book introduced some interesting innovations of applied value to clustering literature." —Journal of Classification "This is a very good, easy-to-read, and practical book. It has many nice features and is highly recommended for students and practitioners in various fields of study." —Technometrics An introduction to the practical application of cluster analysis, this text presents a selection of methods that together can deal with most applications. These methods are chosen for their robustness, consistency, and general applicability. This book discusses various types of data, including interval-scaled and binary variables as well as similarity data, and explains how these can be transformed prior to clustering.

Get valuable insights from your data by building data analysis systems from scratch with R. About This Book A handy guide to take your understanding of data analysis with R to the next level Real-world projects that focus on problems in finance, network analysis, social media, and more From data manipulation to analysis to visualization in R, this book will teach you everything you need to

know about building end-to-end data analysis pipelines using R Who This Book Is For If you are looking for a book that takes you all the way through the practical application of advanced and effective analytics methodologies in R, then this is the book for you. A fundamental understanding of R and the basic concepts of data analysis is all you need to get started with this book. What You Will Learn Build end-to-end predictive analytics systems in R Build an experimental design to gather your own data and conduct analysis Build a recommender system from scratch using different approaches Use and leverage RShiny to build reactive programming applications Build systems for varied domains including market research, network analysis, social media analysis, and more Explore various R Packages such as RShiny, ggplot, recommenderlab, dplyr, and find out how to use them effectively Communicate modeling results using Shiny Dashboards Perform multi-variate time-series analysis prediction, supplemented with sensitivity analysis and risk modeling In Detail R offers a large variety of packages and libraries for fast and accurate data analysis and visualization. As a result, it's one of the most popularly used languages by data scientists and analysts, or anyone who wants to perform data analysis. This book will demonstrate how you can put to use your existing knowledge of data analysis in R to build highly efficient, end-to-end data analysis pipelines without any hassle. You'll start by building a content-based recommendation system, followed by building a project on sentiment analysis with tweets. You'll implement time-series

modeling for anomaly detection, and understand cluster analysis of streaming data. You'll work through projects on performing efficient market data research, building recommendation systems, and analyzing networks accurately, all provided with easy to follow codes. With the help of these real-world projects, you'll get a better understanding of the challenges faced when building data analysis pipelines, and see how you can overcome them without compromising on the efficiency or accuracy of your systems. The book covers some popularly used R packages such as dplyr, ggplot2, RShiny, and others, and includes tips on using them effectively. By the end of this book, you'll have a better understanding of data analysis with R, and be able to put your knowledge to practical use without any hassle. Style and approach This book takes a unique, learn-as-you-do approach, as you build on your understanding of data analysis progressively with each project. This book is designed in a way that implementing each project will empower you with a unique skill set, and enable you to implement the next project more confidently.

Handbook of Cluster Analysis provides a comprehensive and unified account of the main research developments in cluster analysis. Written by active, distinguished researchers in this area, the book helps readers make informed choices of the most suitable clustering approach for their problem and make better use of existing cluster analysis tools. The book is organized according to the traditional core approaches to cluster analysis, from the origins to recent developments. After an overview of approaches and a quick journey through

the history of cluster analysis, the book focuses on the four major approaches to cluster analysis. These approaches include methods for optimizing an objective function that describes how well data is grouped around centroids, dissimilarity-based methods, mixture models and partitioning models, and clustering methods inspired by nonparametric density estimation. The book also describes additional approaches to cluster analysis, including constrained and semi-supervised clustering, and explores other relevant issues, such as evaluating the quality of a cluster. This handbook is accessible to readers from various disciplines, reflecting the interdisciplinary nature of cluster analysis. For those already experienced with cluster analysis, the book offers a broad and structured overview. For newcomers to the field, it presents an introduction to key issues. For researchers who are temporarily or marginally involved with cluster analysis problems, the book gives enough algorithmic and practical details to facilitate working knowledge of specific clustering areas.

Cluster randomization is frequently used in clinical trials for convenience of inter ventional implementation and for reducing the risk of contamination. The opera tional convenience of cluster randomized trials, however, is gained at the expense of reduced analytical power. Compared to individually randomized studies, cluster randomized trials often have a much-reduced power. In this dissertation, I consider ways of enhancing analytical power with historical trial data. Specifically, I introduce a hierarchical Bayesian model that is designed to incorporate available information from previous trials of

the same or similar interventions. Operationally, the amount of information gained from the previous trials is determined by a Kullback-Leibler divergence measure that quantifies the similarity, or lack thereof, between the histor ical and current trial data. More weight is given to the historical data if they more closely resemble the current trial data. Along this line, I examine the Type I error rates and analytical power associated with the proposed method, in comparison with the existing methods without utilizing the ancillary historical information. Similarly, to design a cluster randomized trial, one could estimate the power by simulating trial data and comparing them with the historical data from the published studies. Data analytical and power simulation methods are developed for more general situations of cluster randomized trials, with multiple arms and multiple types of data following the exponential family of distributions. An R package is developed for practical use of the methods in data analysis and trial design.

One-size-fits-all cluster policies have been rightly criticized in the literature. One promising approach is to focus cluster policies on the specific needs of firms depending on the stage of development (emergence, growth, sustainment or decline) their cluster is in. In this highly insightful book, these stage-specific cluster policies are analysed and evaluated. Moreover, several chapters also focus on smart specialization policies to promote regional development by taking into account the emergence and adaptation of clusters and industries. This accessible, practice-oriented and compact text

provides a hands-on introduction to market research. Using the market research process as a framework, it explains how to collect and describe data and presents the most important and frequently used quantitative analysis techniques, such as ANOVA, regression analysis, factor analysis and cluster analysis. The book describes the theoretical choices a market researcher has to make with regard to each technique, discusses how these are converted into actions in IBM SPSS version 22 and how to interpret the output. Each chapter concludes with a case study that illustrates the process using real-world data. A comprehensive Web appendix includes additional analysis techniques, datasets, video files and case studies. Tags in the text allow readers to quickly access Web content with their mobile device. The new edition features: Stronger emphasis on the gathering and analysis of secondary data (e.g., internet and social networking data) New material on data description (e.g., outlier detection and missing value analysis) Improved use of educational elements such as learning objectives, keywords, self-assessment tests, case studies, and much more Streamlined and simplified coverage of the data analysis techniques with more rules-of-thumb Uses IBM SPSS version 22

Cluster Randomised Trials, Second Edition discusses the design, conduct, and analysis of trials that randomise groups of individuals to different treatments. It explores the advantages of cluster randomisation, with special attention given to evaluating the effects of interventions against infectious diseases. Avoiding unnecessary mathematical detail, the book covers basic concepts underlying the use of cluster randomisation, such as direct, indirect, and total

effects. In the time since the publication of the first edition, the use of cluster randomised trials (CRTs) has increased substantially, which is reflected in the updates to this edition. There are greatly expanded sections on randomisation, sample size estimation, and alternative designs, including new material on stepped wedge designs. There is a new section on handling ordinal outcome data, and an appendix with descriptions and/or generating code of the example data sets. Although the book mainly focuses on medical and public health applications, it shows that the rigorous evidence of intervention effects provided by CRTs has the potential to inform public policy in a wide range of other areas. The book encourages readers to apply the methods to their own trials, reproduce the analyses presented, and explore alternative approaches.

Winner of the 2016 De Groot Prize from the International Society for Bayesian Analysis 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.

A cluster randomization trial is one in which intact social units, or clusters of individuals, are randomized to different intervention groups. Trials randomizing clusters have become particularly widespread in the evaluation of non-therapeutic interventions, including lifestyle modification, educational programmes and innovations in the provision of health care. The increasing popularity of this design among health researchers over the past two decades has led to an extensive body of methodology on the subject. This is the first book to present a systematic and united treatment of this topic; it contains distinctive chapters on the history of cluster randomized trials, ethical issues and reporting guidelines.

Balancing simplicity with technical rigour, this practical guide to the statistical techniques essential to research in marketing and related fields, describes each method as well as showing how they are applied. The book is accompanied by two real data sets to replicate examples and with exercises to solve, as well as detailed guidance on the use of appropriate software including: - 750 powerpoint slides with lecture notes and step-by-step guides to run analyses in SPSS (also includes screenshots) - 136 multiple choice questions for tests This is augmented by in-depth discussion of topics including: - Sampling - Data management and statistical packages - Hypothesis testing - Cluster analysis - Structural

equation modelling