<<   >>

<<   >>

13   ISBN   9787040247558

10   ISBN   7040247550

          2008-12

     269

                              PDF

                    http://www.tushu007.com

The first eight years of the twenty-first century has witted the explosion of datacollection, with relatively low costs. Data with curves, images and movies are fre-quently collected in molecular biology, health science, engineering, geology, clima-tology, economics, finance, and humanities. For example, in biomedical research, MRI, fMRI, microarray, and proteomics data are frequently collected for eachsubject, involving hundreds of subjects; in molecular biology, massive sequencingdata are becoming rapidly available; in natural resource discovery and agricul-ture, thousands of high-resolution images are collected; in business and finance, millions of transactions are recorded every day. Frontiers of science, engineering, and humanities differ in the problems of their concerns, but nevertheless share acommon theme: massive or complex data have been collected and new knowledgeneeds to be discovered. Massive data collection and new scientific research havestrong impact on statistical thinking, methodological development, and theoreti-cal studies. They have also challenged traditional statistical theory, methods, andcomputation. Many new insights and phenomena need to be discovered and newstatistical tools need to be developed. With this background, the Center for Statistical Research at the ChineseAcademy of Science initiated the conference series "International Conference onthe Frontiers of Statistics" in 2005. The aim is to provide a focal venue for re-searchers to gather, interact, and present their new research findings, to discussand outline emerging problems in their fields, to lay the groundwork for future col-laborations, and to engage more statistical scientists in China to conduct researchin the frontiers of statistics. After the general conference in 2005, the 2006 Inter-national Conference on the Frontiers of Statistics, held in Changchun, focused onthe topic "Biostatistics and Bioinformatics". The conference attracted many topresearchers in the area and was a great success. However, there are still a lot ofChinese scholars, particularly young researchers and graduate students, who werenot able to attend the conference. This hampers one of the purposes of the con-ference series. However, an Mternative idea was born: inviting active researchersto provide a bird-eye view on the new developments in the frontiers of statistics, on the theme topics of the conference series. This will broaden significantly thebenefits of statistical research, both in China and worldwide. The edited books inthis series aim at promoting statistical research that has high societal impacts andprovide not only a concise overview on the recent developments in the frontiers ofstatistics, but also useful references to the literature at large, leading readers trulyto the frontiers of statistics.

presents an overview of recent developments in biostatistics and bioinformatics. Written by active researchers in these emerging areas, it is intended to give graduate students and new researchers an idea of where the frontiers of biostatistics and bioinformatics are as well as a forum to learn common techniques in use, so that they can advance the fields via developing new techniques and new results. Extensive references are provided so that researchers can follow the threads to learn more comprehensively what the literature is and to conduct their own research. In particulars, the book covers three important and rapidly advancing topics in biostatistics: analysis of survival and longitudinal data, statistical methods for epidemiology.

PrefacePart Analysis of Survival and Longitudinal DataChapter 1 Non- and Semi- Parametric Modeling in Survival Analysis 1 Introduction 2 Cox's type of models 3 Multivariate Cox's type of models 4 Model selection on Cox's models 5 Validating Cox's type of models 6 Transformation models 7 Concluding remarks ReferencesChapter 2 Additive-Accelerated Rate Model for Recurrent Event 1 Introduction 2 Inference procedure and asymptotic properties 3 Assessing additive and accelerated covariates 4 Simulation studies 5 Application 6 Remarks Acknowledgements Appendix ReferencesChapter 3 An Overview on Quadratic Inference Function Approaches for Longitudinal Data 1 Introduction 2 The quadratic inference function approach 3 Penalized quadratic inference function 4 Some applications of QIF 5 Further research and concluding remarks Acknowledgements ReferencesChapter 4 Modeling and Analysis of Spatially Correlated Data 1 Introduction 2 Basic concepts of spatial process 3 Spatial models for non-normal/discrete data 4 Spatial models for censored outcome data 5 Concluding remarks ReferencesPart Statistical Methods for EpidemiologyChapter 5 Study Designs for Biomarker-Based Treatment Selection 1 Introduction 2 Definition of study designs 3 Test of hypotheses and sample size calculation 4 Sample size calculation 5 Numerical comparisons of efficiency 6 Conclusions Acknowledgements Appendix ReferencesChapter 6 Statistical Methods for Analyzing Two-Phase Studies 1 Introduction 2 Two-phase case-control or cross-sectional studies 3 Two-phase designs in cohort studies 4 Conclusions ReferencesPart BioinformaticsChapter 7 Protein Interaction Predictions from Diverse Sources 1 Introduction 2 Data sources useful for protein interaction predictions 3 Domain-based methods 4 Classification methods 5 Complex detection methods 6 Conclusions Acknowledgements ReferencesChapter 8 Regulatory Motif Discovery" From Decoding to Meta-Analysis 1 Introduction 2 A Bayesian approach to motif discovery 3 Discovery of regulatory modules 4 Motif discovery in multiple species 5 Motif learning on ChiP-chip data 6 Using nucleosome positioning information in motif discovery 7 Conclusion ReferencesChapter 9 Analysis of Cancer Genome Alterations Using Singk Nucleotide Polymorphism (SNP) Microarrays 1 Background 2 Loss of heterozygosity analysis using SNP arrays 3 Copy number analysis using SNP arrays 4 High-level analysis using LOH and copy number data 5 Software for cancer alteration analysis using SNP arrays 6 Prospects Acknowledgements ReferencesChapter 10 Analysis of ChiP-chip Data on Genome Tiling Microarrays 1 Background molecular biology 2 A ChiP-chip experiment 3 Data description and analysis 4 Follow-up analysis 5 Conclusion ReferencesSubject IndexAuthor Index

We assume that patients can be divided into two groups based on an assay of a biomarker. This biomarker could be a composite of hundreds of molecular and genetic factors, for example, but in this case we suppose that a cutoff value has been determined that dichotomizes these values. In our example the biomarker is the expression of guanylyl cyclase C (GCC) in the lymph nodes of patients. We assume that we have an estimate of the sensitivity and specificity of the biomarker assay. The variable of patient response is taken to be continuous-valued; it could represent a measure of toxicity to the patient, quality of life, uncensored survival time, or a composite of several measures. In our example we take the endpoint to be three-year disease recurrence. We consider five study designs, each addressing its own set of scientific ques-tions, to study how patients in each marker group fare with each treatment. Al-though consideration of which scientific questions are to be addressed by the study should supersede consideration of necessary sample size, we give efficiency com-parisons here for those cases in which more than one design would be appropriate. One potential goal is to investigate how treatment assignment and patient marker status affect outcome, both separately and interactively. The marker under con-sideration is supposedly predictive: it modifies the treatment effect. We may want to verify its predictive value and to assess its prognostic value, that is, how well it divides patients receiving the same treatment into different risk groups. Each study design addresses different aspects of these overarching goals. This paper is organized as follows: 1. Definition of study designs 2. Test of hypotheses 3. Sample size calculation 4. Numerical comparison of efficiency 5. Conclusions 2 Definition of study designs The individual study designs are as follows. 2.1 Traditional design To assess the safety and efficacy of the novel treatment, the standard design (Fig. 1) is to register patients, then randomize thorn with ratio ~ ~ to receive treatment A or B. We compare the response variable across the two arms of the trial without regard for the marker status of the patients. In our example, we would utilize this design if we wanted only to compare the recurrence rates of colorectal cancer in the two treatment groups independent of each patient's biomarker status.

PDF

:http://www.tushu007.com