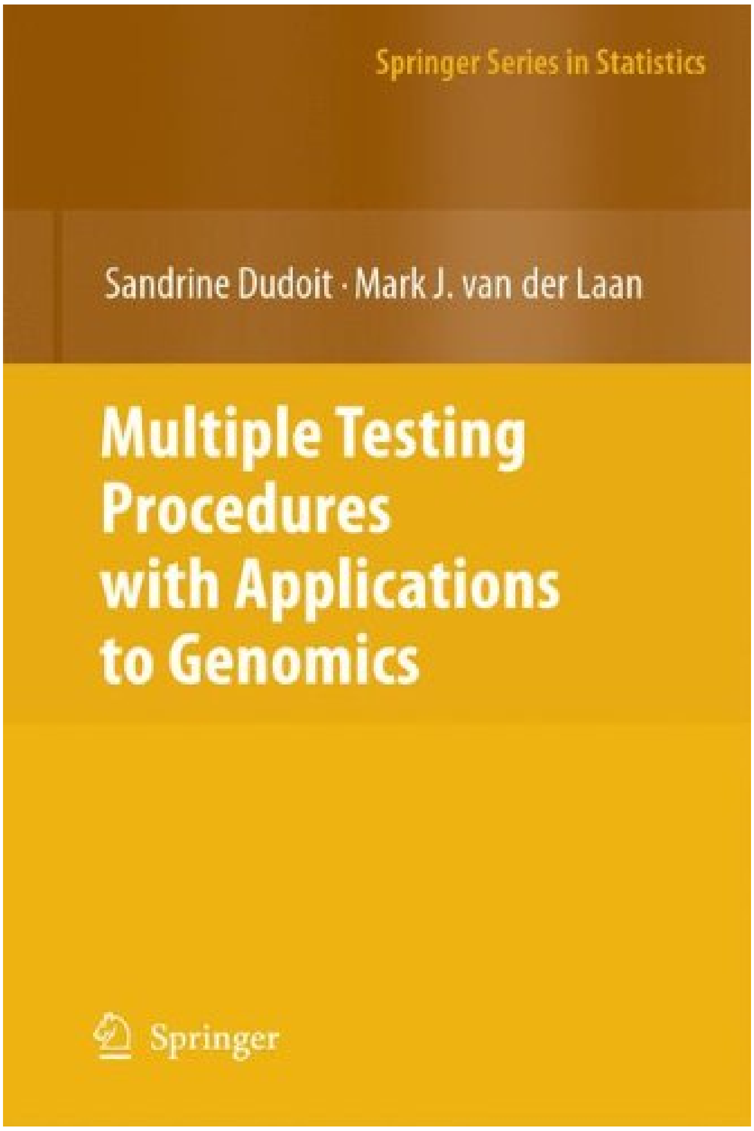
# Multiple Testing Procedures with Applications to Genomics (Springer Series in Statistics)

*Sandrine Dudoit, Mark J. van der Laan*
*DOC | *audiobook | ebooks | Download PDF | ePub*



#3596606 in Books Sandrine Dudoit 2007-12-19 2007-12-19Original language:EnglishPDF # 1 9.25 x 1.42 x 6.10l, 2.21 #File Name: 0387493166590 pagesMultiple Testing Procedures with Applications to Genomics | File size: 32.Mb

**Sandrine Dudoit, Mark J. van der Laan : Multiple Testing Procedures with Applications to Genomics (Springer Series in Statistics)** before purchasing it in order to gage whether or not it would be worth my time, and all praised Multiple Testing Procedures with Applications to Genomics (Springer Series in Statistics):

3 of 3 people found the following review helpful. multiple testing with microarraysBy Michael R. ChernickThe

traditional approach to multiple testing or simultaneous inference was to take a small number of correlated or uncorrelated tests and estimate a family-wise type I error rate that minimizes the the probability of just one type I error out of the whole set whan all the null hypotheses hold. Bounds like Bonferroni or Sidak were sometimes used to as method for constraining the typeI error as they represented upper bounds. Other approaches were to use multivariate methods for tests statistics such as Tukey's least significant difference, Scheffe's method and Dunnett's test. More recently stepdown procedures have become popular in clinical trials but there the multiplicity is usually 5 or less. With the introduction of the bootstrap and advances in computer speed that allowed permutation methods to gain a greater prominence also Westfall and Young came up with a prescription for using resampling to adjust individual p-values for the multiple testing and this was implemented in the SAS procedure MULTTEST and documented both in the SAS manual and the book by Westfall and Young in the mid 1990s. The authors of this text want to extend multiple testing to microarrays where literally thousands of hypothesis are being tested on a single array. Dudoit and van der Laan extend the theory to permit bootstrapping to work in a much broader context where many criteria other than familywise error rate (FWER)are considered including false discovery rate (FDR). They say that for problems involving very high dimensional data an assumption they call subset pivotality does not apply. This assumption is essentially what is needed in the Westfall and Young theory and involves the use of what the authors call a data generating null distribution. To create a method that works for microarray and other high dimensional data the authors base their procedrues onthe joint null distribution of the test statistics rather than the data generating null distributions that all other methods depend on.The book provides a very general theory that generalizes the ideas of resampling based methods to a new framework. The authors intend the book for both statisticians and applied scientists who encounter high-dimensional data in their subject area. The book provides a very detailed and highly theoretical account of multiple testing and may not be suitable for some applied statisticians and scientists. But the ideas are important to all especially in the area of genomics. The authors claim that chapters 4-7 are theoretical chapters that may not be suitable for everyone but they insist that the introductory chapters 1-3 and the applications chapters 8-13 are intended for people with a good biological background but not necessarily a very strong statistical background. I do not share their view about chapters 1-3 which I think would be difficult for anyone lack a graduate level statistics background but I do agree that the applications chapters 8-13 are palatable for the intended audience and is particular interesting for those with knowledge of and interest in the biological sciences.

This book establishes the theoretical foundations of a general methodology for multiple hypothesis testing and discusses its software implementation in R and SAS. These are applied to a range of problems in biomedical and genomic research, including identification of differentially expressed and co-expressed genes in high-throughput gene expression experiments; tests of association between gene expression measures and biological annotation metadata; sequence analysis; and genetic mapping of complex traits using single nucleotide polymorphisms. The procedures are based on a test statistics joint null distribution and provide Type I error control in testing problems involving general data generating distributions, null hypotheses, and test statistics.

From the reviews: "This book summarizes the recent work of Sandrine Dudoit and Mark van der Laan on multiple testing. It proposes a general framework for multiple testing procedures (MTPs) and introduces new concepts . The authors also provide code for reproducing the results of some of the applications. if one is looking for a detailed summary of the latest developments in multiple testing regarding MTPs or in the application of MTPs to biomedical and genomic data, then this book is an excellent reference." (Holger Schwender, Statistical Papers, Vol. 50, 2009) "In the last decade a growing amount of statistical research has been devoted to multiple testing. This book summarizes the recent work on this area. very useful for the applied researcher who would like to understand how to apply multiple testing. a good reference for statisticians interested in a general treatment of multiple testing." (Avner Bar-Hen, Mathematical s, Issue 2009 j)From the Back CoverThis book establishes the theoretical foundations of a general methodology for multiple hypothesis testing and discusses its software implementation in R and SAS. The methods are applied to a range of testing problems in biomedical and genomic research, including the identification of differentially expressed and co-expressed genes in high-throughput gene expression experiments, such as microarray experiments; tests of association between gene expression measures and biological annotation metadata (e.g., Gene Ontology); sequence analysis; and the genetic mapping of complex traits using single nucleotide polymorphisms. The book is aimed at both statisticians interested in multiple testing theory and applied scientists encountering high-dimensional testing problems in their subject matter area. Specifically, the book proposes resampling-based single-step and stepwise multiple testing procedures for controlling a broad class of Type I error rates, defined as tail probabilities and expected values for arbitrary functions of the numbers of Type I errors and rejected hypotheses (e.g., false discovery rate). Unlike existing approaches, the procedures are based on a test statistics joint null distribution and provide Type I error control in testing problems involving general data generating distributions (with arbitrary dependence structures among variables), null hypotheses, and test statistics. The multiple testing results are reported in terms of rejection regions, parameter confidence regions, and adjusted p-values. Sandrine Dudoit is Associate

Professor of Biostatistics and Statistics at the University of California, Berkeley (www.stat.berkeley.edu/~sandrine). Her research and teaching activities concern the development and application of statistical and computational methods for the analysis of high-dimensional biomedical and genomic data. She is a founding core developer of the Bioconductor Project and is an Associate Editor for six journals, including the Annals of Applied Statistics and Statistical Applications in Genetics and Molecular Biology. Mark J. van der Laan is Hsu/Peace Professor of Biostatistics and Statistics at the University of California, Berkeley (www.stat.berkeley.edu/~laan). His research concerns causal inference, adjusting for missing and censored data, and simultaneous estimation and testing based on high-dimensional observational and experimental biomedical and genomic data. He is co-author with James Robins of Unified Methods for Censored Longitudinal Data and Causality (Springer, 2003). He is a recipient of the 2005 COPSS Presidents' and Snedecor Awards and is an active Associate Editor for five journals, including the Annals of Statistics and the International Journal of Biostatistics.