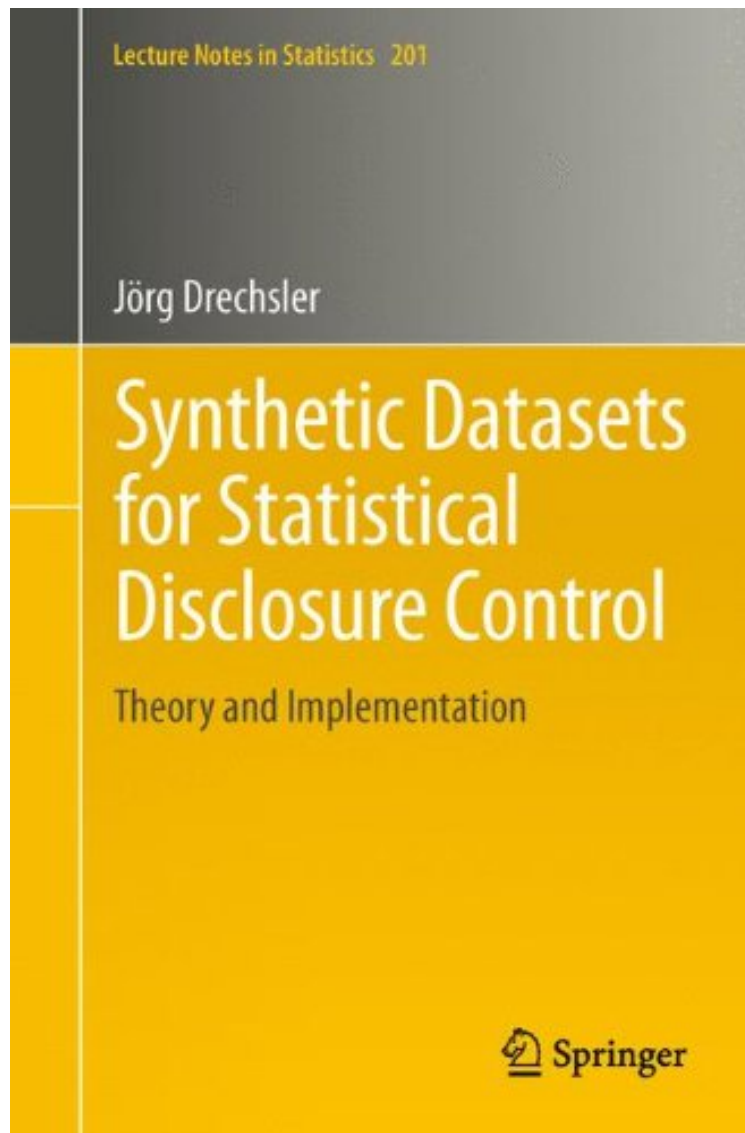


(Download free pdf) Synthetic Datasets for Statistical Disclosure Control: Theory and Implementation
(Lecture Notes in Statistics)

Synthetic Datasets for Statistical Disclosure Control: Theory and Implementation (Lecture Notes in Statistics)

Jörg Drechsler

ebooks | Download PDF | *ePub | DOC | audiobook



 Download

 Read Online

#3807978 in Books J Rg Drechsler 2011-06-29Original language:EnglishPDF # 1 9.25 x .37 x 6.10l, .51
#File Name: 1461403251138 pagesSynthetic Datasets for Statistical Disclosure Control Theory and
Implementation Lecture Notes in Statistics | File size: 43.Mb

Jrg Drechsler : Synthetic Datasets for Statistical Disclosure Control: Theory and Implementation (Lecture Notes in Statistics) before purchasing it in order to gage whether or not it would be worth my time, and all praised

The aim of this book is to give the reader a detailed introduction to the different approaches to generating multiply imputed synthetic datasets. It describes all approaches that have been developed so far, provides a brief history of synthetic datasets, and gives useful hints on how to deal with real data problems like nonresponse, skip patterns, or logical constraints. Each chapter is dedicated to one approach, first describing the general concept followed by a detailed application to a real dataset providing useful guidelines on how to implement the theory in practice. The discussed multiple imputation approaches include imputation for nonresponse, generating fully synthetic datasets, generating partially synthetic datasets, generating synthetic datasets when the original data is subject to nonresponse, and a two-stage imputation approach that helps to better address the omnipresent trade-off between analytical validity and the risk of disclosure. The book concludes with a glimpse into the future of synthetic datasets, discussing the potential benefits and possible obstacles of the approach and ways to address the concerns of data users and their understandable discomfort with using data that doesn't consist only of the originally collected values. The book is intended for researchers and practitioners alike. It helps the researcher to find the state of the art in synthetic data summarized in one book with full reference to all relevant papers on the topic. But it is also useful for the practitioner at the statistical agency who is considering the synthetic data approach for data dissemination in the future and wants to get familiar with the topic.

From the reviews: This book explores the generation of synthetic data with the same statistical properties as the original data, but containing different records. It also addresses complications which arise with real data, such as missing values. The book provides an excellent overview of the area, and would make ideal reading for anyone new to the issues, as well as serving as a good source book describing technical details. (David J. Hand, *International Statistical*, Vol. 80 (3), 2012)

From the Back Cover: The aim of this book is to give the reader a detailed introduction to the different approaches to generating multiply imputed synthetic datasets. It describes all approaches that have been developed so far, provides a brief history of synthetic datasets, and gives useful hints on how to deal with real data problems like nonresponse, skip patterns, or logical constraints. Each chapter is dedicated to one approach, first describing the general concept followed by a detailed application to a real dataset providing useful guidelines on how to implement the theory in practice. The discussed multiple imputation approaches include imputation for nonresponse, generating fully synthetic datasets, generating partially synthetic datasets, generating synthetic datasets when the original data is subject to nonresponse, and a two-stage imputation approach that helps to better address the omnipresent trade-off between analytical validity and the risk of disclosure. The book concludes with a glimpse into the future of synthetic datasets, discussing the potential benefits and possible obstacles of the approach and ways to address the concerns of data users and their understandable discomfort with using data that doesn't consist only of the originally collected values. The book is intended for researchers and practitioners alike. It helps the researcher to find the state of the art in synthetic data summarized in one book with full reference to all relevant papers on the topic. But it is also useful for the practitioner at the statistical agency who is considering the synthetic data approach for data dissemination in the future and wants to get familiar with the topic.

About the Author: Jrg Drechsler is a Research Scientist at the German Institute for Employment Research, Department for Statistical Methods. His main areas of research involve statistical disclosure control and imputation with published papers in *JASA*, *Statistica Sinica*, *JOS*, and *Survey Methodology*.