DATA MANAGEMENT ESSENTIALS USING SAS AND JMP

SAS programming is a creative and iterative process designed to empower you to make the most of your organization's data. This friendly guide provides you with a repertoire of essential SAS tools for data management, whether you're a new or an infrequent user. Most useful to students and programmers with little or no SAS experience, it takes a no-frills, hands-on tutorial approach to getting started with the software.

You'll find immediate guidance in navigating, exploring, visualizing, cleaning, formatting, and reporting on data using SAS and JMP. Step-by-step demonstrations, screenshots, handy tips, and practical exercises with solutions equip you to explore, interpret, process, and summarize data independently, efficiently, and effectively.

Julie Kezik is a biostatistician at the Yale Center for Perinatal, Pediatric, and Environmental Epidemiology. Her primary research interests are assessing the effects of indoor and outdoor air pollution on at-risk populations, and she currently focuses on providing statistical analysis and research support for epidemiological studies of environmental exposures and early childhood health outcomes. Kezik's current work uses a combination of measured data, traffic information, and health outcomes to help create interventions that will improve health.

Melissa Hill is a clinical programmer at Cd3 Inc. where she uses SAS to perform and support the design and programming of clinical data structures related to drug development. Prior to her position at Cd3 Inc., she worked as an epidemiologist at the Yale Center for Perinatal, Pediatric, and Environmental Epidemiology. During that time, Hill used SAS to support her various roles at the CPPEE including programmer analyst, field study coordinator, and research associate. She enjoys sharing her diverse SAS experience with other members of her team and developing new ways to harness the broad range of tools that SAS provides.

DATA MANAGEMENT ESSENTIALS USING SAS AND JMP

JULIE KEZIK, MS

MELISSA HILL, MPH

CAMBRIDGE UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment, a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107535039

First published 2016

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging-in-Publication data
Names: Kezik, Julie, 1982– author. | Hill, Melissa E., 1982– author.
Title: Data management essentials using SAS and JMP / Julie Kezik, Yale School of Public Health, Melissa Hill, Yale University Center for Perinatal Pediatric and Environmental Epidemiology.
Description: New York : Cambridge University Press, 2016. | Includes bibliographical references.
Identifiers: LCCN 2015043490
Subjects: LCSH: Database management. | Data structures (Computer science) | SAS (Computer file) | JMP (Computer file)
Classification: LCC QA76.9.D3 K358984 2016 | DDC 005.74–dc23
LC record available at http://lccn.loc.gov/2015043490

ISBN 978-1-107-11456-2 Hardback
ISBN 978-1-107-53503-9 Paperback

Contents

Acknowledgments

We would like to thank our friends and colleagues at the Yale University Center for Perinatal, Pediatric, and Environmental Epidemiology for their support and encouragement.

Another very special expression of gratitude to our families for countless hours of childcare.

About This Book

In research groups around the world, SAS is used not only by statisticians and investigators for data analysis but by programmers and data managers to handle seemingly endless libraries of priceless data. These data management teams often include support staff who have been selected and hired for their attention to detail and patience with meticulous tasks, but who are not necessarily fluent in SAS. The typical solution is for more advanced users to do excessive and simplistic programming to provide the output necessary for whatever task the assistant will be handling. This cycle creates extra work for advanced users and limits the independent effectiveness of the support team – a frustrating arrangement for all parties involved.

In the face of this conundrum we sought a training program for our support team. We found that no training program existed that met our specific needs; available resources were either too costly, too time consuming, or too statistically driven. Eventually we developed and initiated our own basic SAS training program for our programming and research assistants. Spending some structured time with employees while they explored SAS was the most economical way to teach basic users the skills they needed to complete daily tasks. Since training our own staff, we have experienced increased productivity by an empowered support team. This book is a result of that successful endeavour, which inspired us to share our curriculum with other groups who are undoubtedly faced with the same challenge.

SAS programming is a creative and iterative process designed to empower the user. The purpose of this text is not to instruct users on how to complete specific tasks, but to provide a toolkit of essentials for new and infrequent users. When used appropriately, this book will enable these users to explore, interpret, process, and summarize data independently.

How to Use This Book

This book can be used from cover to cover as a hands-on training manual or simply as a desk reference. The content is directed at first-time or infrequent users who seek immediate applicability in order to navigate, clean, and report data. In an effort to truly teach the most basic SAS skills essential to data management, this text uses a multitude of examples and screenshots to walk the reader through step-by-step instructions for executing commonly used techniques and procedures. Beginning with Chapter 2, there is a 'Test Your Skills' section with practice tasks and full solution sets. You will find that in SAS there is more than one way to accomplish many of the tasks; the solutions provided should in no way be perceived as exhaustive. All of the examples and practice tasks are based on datasets found in the sashelp library or created by you, the user, and require no additional software or downloading. The versions of software used for examples include SAS 9.4, SAS Enterprise Guide 4.3, and JMP Pro 10.