Exercises and Projects for The Little SAS® Book

A Programming Approach

a primer

SIXTH EDITION

Rebecca A. Ottesen
Lora D. Delwiche
Susan J. Slaughter

The correct bibliographic citation for this manual is as follows: Ottesen, Rebecca A., Lora D. Delwiche, and Susan J. Slaughter. 2020. *Exercises and Projects for The Little SAS® Book, Sixth Edition*. Cary, NC: SAS Institute Inc.

Exercises and Projects for The Little SAS® Book, Sixth Edition

Copyright © 2020, SAS Institute Inc., Cary, NC, USA

ISBN 978-1-64295-617-7 (Hard cover)
ISBN 978-1-64295-284-1 (Paperback)
ISBN 978-1-64295-524-8 (Web PDF)
ISBN 978-1-64295-525-5 (epub)
ISBN 978-1-64295-526-2 (Kindle)

All Rights Reserved. Produced in the United States of America.

For a hard copy book: No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

For a web download or e-book: Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

The scanning, uploading, and distribution of this book via the Internet or any other means without the permission of the publisher is illegal and punishable by law. Please purchase only authorized electronic editions and do not participate in or encourage electronic piracy of copyrighted materials. Your support of others' rights is appreciated.

U.S. Government License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication, or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a), and DFAR 227.7202-4, and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513-2414

January 2020

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to **http://support.sas.com/thirdpartylicenses**.

Contents

About This Book vii

About The Authors xi

Acknowledgments xiii

Exercises and Projects

Chapter 1 Getting Started Using SAS Software 1

Multiple Choice **1**

Short Answer **4**

Programming Exercises **5**

Chapter 2 Accessing Your Data 7

Multiple Choice **7**

Short Answer **12**

Programming Exercises **14**

Chapter 3 Working with Your Data 21

Multiple Choice **21**

Short Answer **27**

Programming Exercises **30**

Chapter 4 Sorting, Printing, and Summarizing Your Data 35

Multiple Choice **35**

Short Answer **40**

Programming Exercises **41**

Chapter 5 Enhancing Your Output with ODS 45

Multiple Choice **45**

Short Answer **48**

Programming Exercises **49**

Chapter 6 Modifying and Combining SAS Data Sets 53
Multiple Choice **53**
Short Answer **59**
Programming Exercises **60**
Chapter 7 Writing Flexible Code with the SAS Macro Facility 65
Multiple Choice **65**
Short Answer **68**
Programming Exercises **69**
Chapter 8 Visualizing Your Data 73
Multiple Choice **73**
Short Answer **76**
Programming Exercises **77**
Chapter 9 Using Basic Statistical Procedures 81
Multiple Choice **81**
Short Answer **84**
Programming Exercises **85**
Chapter 10 Exporting Your Data 89
Multiple Choice **89**
Short Answer **91**
Programming Exercises **91**
Chapter 11 Debugging Your SAS Programs 93
Multiple Choice **93**
Short Answer **97**
Programming Exercises **98**
Projects 103

Selected Solutions 115

Chapter 1—Getting Started Using SAS Software **116**

Chapter 2—Accessing Your Data **117**

Chapter 3—Working with Your Data **120**

Chapter 4—Sorting, Printing, and Summarizing Your Data **122**

Chapter 5—Enhancing Your Output with ODS **124**

Chapter 6—Modifying and Combining SAS Data Sets **126**

Chapter 7—Writing Flexible Code with the SAS Macro Facility **127**

Chapter 8—Visualizing Your Data **129**

Chapter 9—Using Basic Statistical Procedures **130**

Chapter 10—Exporting Your Data **131**

Chapter 11—Debugging Your SAS Programs **132**

About This Book

Who should use this book This book was written for anyone who wants hands-on experience while learning or practicing the basics of SAS programming. The exercises and projects in this book are designed so that they can be used in a classroom setting or by an individual reader working alone.

Content This book consists of exercises (with selected solutions) and projects. Each chapter in this book covers the same material as the corresponding chapter in *The Little SAS Book, Sixth Edition*. The content in each chapter builds on previous chapters, so we do not expect a reader to use a technique in one chapter that is introduced in a later chapter.

> **Exercises** There are three types of exercises in this book: multiple choice, short answer, and programming. Each type of exercise is designed to improve the comprehension of topics and syntax, and to develop SAS coding skills with realistic data sets.
>
> **Solutions** Solutions are provided for odd-numbered multiple choice and short answer questions. Solutions are not provided for programming exercises, but hints are provided to help give readers direction.
>
> **Projects** The projects in the final chapter are designed to serve as a comprehensive capstone for the programming skills developed throughout the book. Each project gives readers a chance to synthesize the material learned from various chapters. Completing a project will take several days of intense thought and coding, and will result in a final product that could serve as evidence of SAS proficiency to a future employer or course instructor. Projects can be completed by a group or an individual.

Typographical Conventions This book uses the same typographical conventions as *The Little SAS Book*. SAS doesn't care whether your programs are written in uppercase or lowercase, so you can write your programs any way you want. In this book, we have used uppercase and lowercase to tell you something. All SAS keywords appear in uppercase letters. A keyword is an instruction to SAS and must be spelled correctly. Anything that programmers make up such as variable names, names for SAS data sets, comments, or titles are generally shown in lowercase or mixed-case letters.

Data Sets We have tried to include programming exercises with data sets that are somewhat large, unruly, and messy. We hope that this will help our readers develop their skills more fully. Some of the exercises are based on data that come directly from public sources, while others have been designed to mimic real-life scenarios.

The data sets for exercises can be accessed via web pages for any of the authors of this book at the support site for SAS, support.sas.com/publishing/authors (select the name of the author;

then, look for the cover thumbnail of this book and select Exercise Data). The data sets are organized by chapter, and they are also available in a single, downloadable ZIP file. If you are unable to access the data through the website, send an email to saspress@sas.com.

The projects in the final chapter use real data. We give instructions for locating and downloading the data from primary sources so that you will have the most up-to-date and authentic data possible.

How to use this book The exercises are organized in such a way as to help readers develop useful and sensible programming habits. Readers are encouraged to examine all raw data files and SAS data sets they will be accessing before writing any code.

In many of the programming exercises, we do not explicitly ask the reader to use a particular function or procedure. This was intentional in order to provide users with the opportunity to develop code based on what they learned from the topics that are covered in the corresponding chapter. This means that the code written by one person may not be the same as that written by another person, even though they both may arrive at the same answer.

While the material presented in the book includes exercises for certain graphical and statistical analyses, the purpose of these chapters is merely to introduce the basic syntax and concepts, and not the fine details of the analysis itself. This book may be used in a statistics course as a programming supplement. However, it should not be considered a statistical textbook.

Compatibility with SAS software This book was developed using SAS 9.4, but the exercises are compatible with SAS 9.3 and later. Most exercises require only Base SAS, but a few in Chapter 9 require SAS/STAT, and a few in Chapters 2 and 10 require SAS/ACCESS Interface to PC Files. The exercises in this book were designed to be programmed in the SAS windowing environment, SAS Enterprise Guide, SAS Studio, SAS OnDemand for Academics, or the SAS University Edition through SAS Analytics U. We have noted the few instances when an exercise is specific to the SAS windowing environment. Otherwise, the exercise can be considered environment-independent. Path references in this book are, however, specific to Microsoft Windows, and may need to be adjusted for other environments such as UNIX, Linux, or z/OS. (SAS University Edition runs in a Linux environment even if you are using a Windows PC.)

For Instructors Instructors may obtain a complete solutions manual by completing the form found at the "Instructors Solutions" link on the web pages for any of the authors at support.sas.com/publishing/authors, and providing the appropriate credentials. The solutions given for the programming exercises are not the only ones possible. Many of the exercises can be solved in multiple ways. Instructors may choose to share a snippet of output or a graph if they think their students need a little extra help to complete an exercise. The section references that accompany the instructors' solutions refer to the sections in *The Little SAS Book, Sixth Edition* where the material is covered. These references are listed only for the corresponding

chapter of the exercise even though the answer may use subject matter from a previous chapter.

We encourage instructors to add their own follow-up questions to any of the exercises for class assignments. In Chapter 9, "Using Basic Statistical Procedures," we have supplied many extra variables in the data sets so that instructors can append additional questions according to the content covered in their course.

For the projects in the final chapter, instructors may choose to download and distribute the data for the class to use. Instructors may clean the data for the class as they see fit, or leave the data manipulation to the students instead.

Additional Help Although this book illustrates many programming tasks and analyses regularly performed in businesses across industries, questions specific to your aims and issues may arise. To fully support you, SAS Institute and SAS Press offer you the following resources:

- For questions about topics covered in this book, contact the authors through SAS Press at saspress@sas.com.
- For questions about topics in or beyond the scope of this book, post queries to the relevant SAS Support Communities at communities.sas.com.
- To search the SAS user Knowledge Base or contact SAS Customer Support you can visit support.sas.com.

About The Authors

Rebecca A. Ottesen first learned SAS as a student at California Polytechnic State University, San Luis Obispo, where she now teaches for the Statistics Department. As a biostatistician for the City of Hope, Rebecca uses every opportunity to incorporate her research and programming experience into the coursework for her Cal Poly students.

Lora D. Delwiche enjoys teaching people about SAS software and likes solving challenging problems using SAS. She has spent most of her career at the University of California, Davis, using SAS in support of teaching and research.

Susan J. Slaughter discovered SAS software in graduate school at North Carolina State University. Since then, she has used SAS in a variety of business and academic settings managing large databases, teaching SAS software classes, and helping other SAS programmers. She now works as a consultant through her company, Avocet Solutions.

Learn more about these authors by visiting their author pages, where you can download free book excerpts, access example data, read the latest reviews, get updates, and more:

support.sas.com/ottesen
support.sas.com/delwiche
support.sas.com/slaughter

SAMPLE EXERCISES

Multiple Choice

1. The raw data file called Class.dat contains three test scores for each of two students in a class. If you submit the following SAS program, what will be the value of the variable represented by p(i) for the first observation after the second time through the iterative DO group?

```
----+----1----+----2----+----3
222 Jimmy 95  85  75
333 Ulric 90  80  70

DATA score;
  INFILE 'c:\MyRawData\Class.dat';
  INPUT ID Name $ Test1 Test2 Test3;
  ARRAY t(3) Test1 - Test3;
  ARRAY p(3) Prop1 - Prop3;
  DO i = 1 TO 3;
    p(i) = t(i) / 100;
  END;
  Total = SUM(Test1 - Test3);
RUN;
```

a. 0.85
b. 0.80
c. 0.75
d. 0.70

2. What is the purpose of using the MAXDEC = *n* option in a PROC MEANS statement?

a. To maximize the computing power by a factor of n
b. To limit the numeric output to n decimal places
c. To use a maximum of n observations in the calculation
d. None of the above

3. Suppose that the following code is submitted to assign a libref name to a macro variable called &lib. Which PROC PRINT will create output for the SAS data set called SASHELP.CARS?

```
%LET lib = sashelp;
```

a. `PROC PRINT DATA = &lib.cars; RUN;`
b. `PROC PRINT DATA = &lib..cars; RUN;`
c. `PROC PRINT DATA = "&lib.cars"; RUN;`
d. `PROC PRINT DATA = "&lib..cars"; RUN;`

4. A CLASS statement is required by which procedure?

a. MEANS
b. TTEST
c. REG
d. ANOVA

Short Answer

5. Examine the following raw data, which contain a patient ID and group designation (A, B, or C) with multiple observations per line. Write the SAS statements that will read the data into variables named ID and Group using a line-hold specifier, and then keep only those patients in groups A and C.

```
----+----1----+----2----+----3
4165 A 2255 B 3312 C 5689 C
1287 A 5454 A 6672 C 8521 B
8936 C 5764 B
```

6. Suppose you have a variable called Gender that stores a value of 1 for males and a 2 for females. Discuss why creating a user-defined format to print a 1 as 'Male' and a 2 as 'Female' may be better than creating a new character variable in a DATA step and then printing the character variable.

7. Describe any advantages of using PROC SQL to join SAS data sets or to add summary statistics compared to the traditional SAS programming methods that use a MERGE statement in a DATA step.

Programming Exercises

8. The United States Geological Survey provides data on earthquakes of historical interest. The SAS data set called EARTHQUAKES contains data about earthquakes

with a magnitude greater than 2.5 in the United States and its territories. The variables are year, month, day, state, and magnitude.

a. Examine this SAS data set including the variable labels and attributes. Create a scatter plot of year and magnitude for earthquakes that occurred in the year 2000 and beyond.
b. Overlay a time series plot of the mean magnitude for each year on the same graph as part a). This line should appear in red on your graph.
c. Include a legend for your graph that labels the time series plot as Mean. The legend should have no border and appear on the bottom right side of the plot.
d. Earthquakes are classified by their magnitude. Overlay reference lines on your graph for light, moderate, strong, major, and great earthquakes defined at magnitudes of 4.0, 5.0, 6.0, 7.0, and 8.0, respectively. These lines should be labeled, dashed, and have 50% transparency.
e. Make sure that all years appear on the X axis of your graph.

9. The local public transportation company would like to conduct a study of the impact of three possible plans on a certain popular bus route in town. Plan one consists of traffic lights timed in sync via a special remote transmitter on the bus that interacts with the traffic signal. Plan two consists of shutting down a regular traffic lane to become an express lane for buses only. Plan three is the current normal bus route with no changes. To assess the impact of a change, the company implemented plan one for a month, and then implemented plan two for a different month. All three plans were measured during summer months to ensure consistency in weather. The time (in minutes) for a bus to complete its route was measured for randomly selected weekdays and weekends, and these data are stored in the SAS data set called BUS.

a. Examine this SAS data set including the variable labels and attributes. Calculate the mean, median, and standard deviation travel time for each plan. Present these descriptive statistics to one decimal place.
b. Test to see whether there is any difference between the mean times for any of the plans, and compare the resulting p-value to alpha = 0.05.
c. Add pairwise comparisons of group means to the analysis from part b).
d. In a comment in your program, discuss your findings. Which plan should the city adopt?

Selected Solutions

1. A
3. B
5.
```
INPUT ID Group $ @@;
IF Group = 'B' THEN DELETE;
```
7. Using PROC SQL allows you to combine data in one step. The use of PROC SORT for a merge can be avoided because PROC SQL creates a Cartesian product. Additionally, PROC SQL does not require the key matching variables to have the same name while merging in a DATA step does. Adding summary statistics to a data set can be accomplished in the same PROC SQL step with a summary function in the SELECT clause, and possibly a GROUP BY clause, which avoids an additional procedure such as PROC MEANS.
8. *Hint:* This exercise will produce only one graphic. Create a variable for average magnitude that is combined with the original data set.
9. *Hint:* Consider modifying the layout of the data set so that you can conduct the appropriate hypothesis tests on route time for the three plans.

Ready to take your SAS® and JMP® skills up a notch?

Be among the first to know about new books, special events, and exclusive discounts.
support.sas.com/newbooks

Share your expertise. Write a book with SAS.
support.sas.com/publish

sas.com/books
for additional books and resources.

sas
THE POWER TO KNOW®

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.
Other brand and product names are trademarks of their respective companies. © 2017 SAS Institute Inc. All rights reserved. M1588358 US.0217