

# Exercises and Projects for

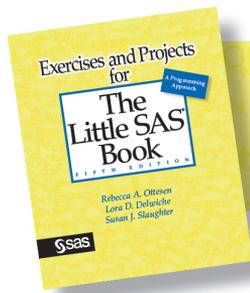
A Programming  
Approach

# The Little SAS<sup>®</sup> Book

F I F T H E D I T I O N

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

sas<sup>®</sup>



From *Exercises and Projects for The Little SAS® Book, Fifth Edition*. Full book available for purchase [here](#).

## Contents

**About the 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 Getting Your Data into SAS 7**

Multiple Choice 7

Short Answer 12

Programming Exercises 14

**Chapter 3 Working with Your Data 21**

Multiple Choice 21

Short Answer 25

Programming Exercises 27

**Chapter 4 Sorting, Printing, and Summarizing Your Data 33**

Multiple Choice 33

Short Answer 38

Programming Exercises 39

**Chapter 5 Enhancing Your Output with ODS 43**

Multiple Choice 43

Short Answer 46

Programming Exercises 47

**Chapter 6 Modifying and Combining SAS Data Sets 51**

Multiple Choice 51

Short Answer 57

Programming Exercises 58

**Chapter 7 Writing Flexible Code with the SAS Macro Facility 63**

Multiple Choice 63

Short Answer 65

Programming Exercises 66

**Chapter 8 Visualizing Your Data 71**

Multiple Choice 71

Short Answer 74

Programming Exercises 74

**Chapter 9 Using Basic Statistical Procedures 79**

Multiple Choice 79

Short Answer 82

Programming Exercises 83

**Chapter 10 Exporting Your Data 87**

Multiple Choice 87

Short Answer 89

Programming Exercises 89

**Chapter 11 Debugging Your SAS Programs 91**

Multiple Choice 91

Short Answer 95

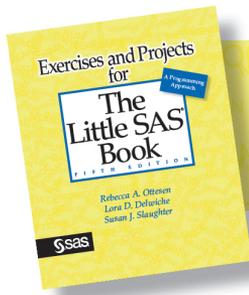
Programming Exercises 96

**Projects 101**

**Selected Solutions 113**

- Chapter 1—Getting Started Using SAS Software 114
- Chapter 2—Getting Your Data into SAS 115
- Chapter 3—Working with Your Data 118
- Chapter 4—Sorting, Printing, and Summarizing Your Data 120
- Chapter 5—Enhancing Your Output with ODS 122
- Chapter 6—Modifying and Combining SAS Data Sets 123
- Chapter 7—Writing Flexible Code with the SAS Macro Facility 125
- Chapter 8—Visualizing Your Data 127
- Chapter 9—Using Basic Statistical Procedures 128
- Chapter 10—Exporting Your Data 129
- Chapter 11—Debugging Your SAS Programs 130

From *Exercises and Projects for The Little SAS® Book, Fifth Edition*, by Rebecca A. Ottesen, Lora D. Delwiche, and Susan J. Slaughter. Copyright © 2015, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.



From *Exercises and Projects for The Little SAS® Book, Fifth Edition*. Full book available for purchase [here](#).

## 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, Fifth 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, <http://support.sas.com/publishing/authors> (select the name of the author; then, look for the cover thumbnail of this book and select Example 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 e-mail to [saspress@sas.com](mailto: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. In Chapter 2, “Getting Your Data into SAS,” readers are encouraged to view the raw data files with a simple editor such as WordPad. In later chapters, readers are encouraged to examine the SAS data sets using tools such as PROC CONTENTS, Viewtable, the Data Grid, or the Properties window or tab.

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, 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 the SAS windowing environment, and may need to be adjusted for other environments such as the SAS University Edition.

**For Instructors** Instructors may obtain a complete solutions manual by completing the form on the web pages for any of the authors of this book at the support site for SAS, <http://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, Fifth 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 can 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 help resources:

- For questions about topics covered in this book, contact the authors through SAS Press:
  - Send questions by email to [saspress@sas.com](mailto:saspress@sas.com); include the book title in your correspondence.
  - Submit feedback on the author’s page at [http://support.sas.com/author\\_feedback](http://support.sas.com/author_feedback).
- For questions about topics in or beyond the scope of this book, post queries to the relevant SAS Support Communities at <https://communities.sas.com/welcome>.
- SAS Institute maintains a comprehensive website with up-to-date information. One page that is particularly useful to both the novice and the seasoned SAS user is its Knowledge Base. Search for relevant notes in the “Samples and SAS Notes” section of the Knowledge Base at <http://support.sas.com/resources>.
- Registered SAS users or their organizations can access SAS Customer Support at <http://support.sas.com>. Here you can pose specific questions to SAS Customer Support; under Support, click Submit a Problem. You will need to provide an email address to which replies can be sent, identify your organization, and

provide a customer site number or license information. This information can be found in your SAS logs.

**Keep in Touch** We look forward to hearing from you. We invite questions, comments, and concerns. If you want to contact us about a specific book, please include the book title in your correspondence.

### **Contact the Authors through SAS Press**

By email: [saspress@sas.com](mailto:saspress@sas.com)

Via the Web: [http://support.sas.com/author\\_feedback](http://support.sas.com/author_feedback)

**Purchase SAS Books** For a complete list of books available through SAS, visit [sas.com/store/books](http://sas.com/store/books).

Phone: 1-800-727-0025

Email: [sasbook@sas.com](mailto:sasbook@sas.com)

**Subscribe to the SAS Training and Book Report** Receive up-to-date information about SAS training, certification, and publications via email by subscribing to the SAS Training & Book Report monthly eNewsletter. Read the archives and subscribe today at <http://support.sas.com/community/newsletters/training>.

**Publish with SAS** SAS is recruiting authors! Are you interested in writing a book? Visit <http://support.sas.com/saspress> for more information.

# 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 DO loop?

```
-----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. 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;
3. A CLASS statement is required by which procedure?

## 2 Exercises and Projects for *The Little SAS Book Fifth Edition*

- a. MEANS
  - b. TTEST
  - c. REG
  - d. ANOVA
4. The following note could be prevented by using what function?

```
NOTE: Character values have been converted to
      numeric values at the places given by:
              (Line):(Column).
              789:3
```

- a. INPUT
- b. PUT
- c. INFILE
- d. FILE

### 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. In the following code, the observations will have a value of missing for the variable HeightCM. Explain why, and how to fix this problem.

```
DATA femaleheight;
  SET height;
  IF Gender = 'Female' THEN OUTPUT;
  HeightCM = HeightIN * 2.54;
RUN;
```

### 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

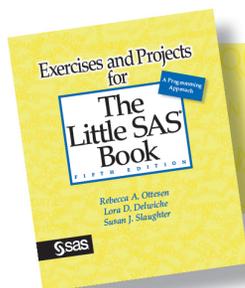
with a magnitude greater than 2.5 in the United States and its territories. The variables are year, month, day, state, and magnitude.

- 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 if 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. D
5. INPUT ID Group \$ @@;  
IF Group = 'B' THEN DELETE;
7. SAS executes statements in the DATA step in the order in which they appear. Therefore, it will output the observations to the SAS data set called FEMALEHEIGHT before it carries out the calculation for the HeightCM variable. The solution would be to swap the statements so that the calculation comes before the OUTPUT statement. Alternatively, you could use a subsetting IF statement instead of an OUTPUT statement. A subsetting IF statement would not override the implied OUTPUT statement at the end of the DATA step.
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.

From *Exercises and Projects for The Little SAS® Book, Fifth Edition*, by Rebecca A. Ottesen, Lora D. Delwiche, and Susan J. Slaughter. Copyright © 2015, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.



From *Exercises and Projects for The Little SAS® Book, Fifth Edition*. Full book available for purchase [here](#).

## About These 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.



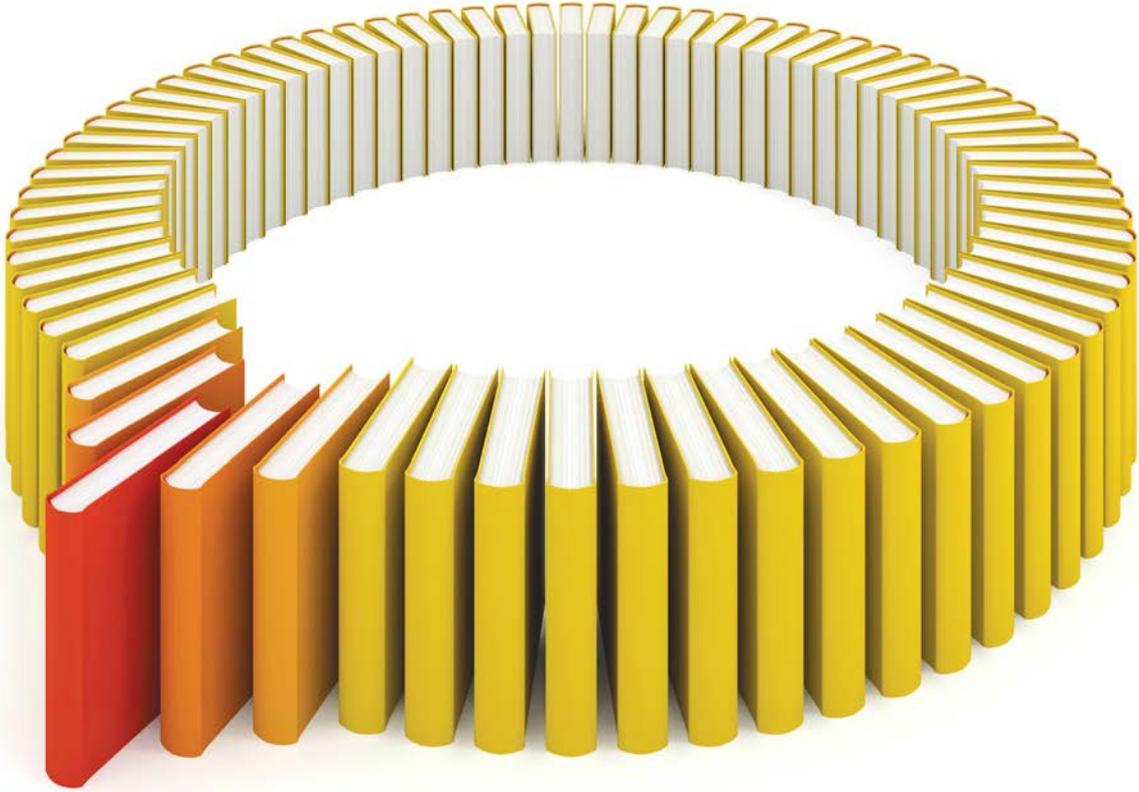
**Susan J. Slaughter** discovered SAS software in graduate school at North Carolina State University. Since then, she has used SAS in a variety of business and academic settings managing large databases, teaching SAS software classes, and helping other SAS programmers. She now works as a consultant through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

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

<http://support.sas.com/publishing/authors/ottesen.html>

<http://support.sas.com/publishing/authors/delwiche.html>

<http://support.sas.com/publishing/authors/slaughter.html>



# Gain Greater Insight into Your SAS<sup>®</sup> Software with SAS Books.

Discover all that you need on your journey to knowledge and empowerment.

 [support.sas.com/bookstore](https://support.sas.com/bookstore)  
for additional books and resources.

  
THE POWER TO KNOW.®