1 Why Do You Need to Learn SAS for Data Analysis?

OBJECTIVE

This chapter presents convincing reasons why readers need the knowledge and skills of SAS for data analysis and file management and how this book can help them acquire these skills.

1.1 Seven Reasons to Learn SAS for Data Analysis

There are at least seven reasons why you should learn SAS. If you don't find any of these reasons convincing, you can stop right here and return this book for a refund—but I doubt this will happen!

First, *knowing SAS is an asset in many job markets*. Many ads ask for SAS experience. The following Web addresses, retrieved in December 2007, represent a handful of Web sites devoted specifically to SAS jobs. Many of these sites include a salary analysis and claim to be one of the fastest growing sites for SAS professionals.

www.sas-jobs.com

http://sas-jobs.dice.com

www.jobster.com/find/US/jobs/for/sas

www.sconsig.com

www.statisticsjobs.com

www.newtyne.com/sas_resources/index.html

www.itjobswatch.co.uk/jobs/uk/sas.do

www.globalstatements.com/sas/jobs/index.html

www.icrunchdata.com

As a matter of fact, a Google® search with the keyword "SAS jobs" will turn up many more Web sites with current job listings that require a variety of SAS expertise. Try it and you will be impressed.

Second, *SAS can read data files created by other statistical packages*. SAS allows data files created by SPSS®, Excel®, Minitab®, Stata®, Systat®, and others to be incorporated into a SAS program directly or through file conversion software. Thus, to experienced users of these statistical packages, SAS presents no threat at all because it is possible to convert data files created by these packages into a SAS file format.

Third, *various data formats can be imported into SAS with relative ease*. In other words, learning SAS will not make you abandon data formats you previously mastered or managed. These formats include those generated and supported by database software such as Oracle®.

Fourth, *SAS is versatile and powerful enough to meet your needs in data analyses*. SAS is flexible, with a variety of input and output formats and numerous procedures for descriptive, inferential, and forecasting types of statistical analyses. Because the SAS System is an integrated system with similar architecture shared by modules or products, once you master one module, you can easily transfer the knowledge to other modules.

Fifth, *there is a worldwide electronic network of SAS users accessible at your fingertips*. This network consists of SAS users who subscribe to a Listserv. Subscribers to this list can converse with each other on the Internet. For details, refer to Appendix A.

Sixth, *the SAS users' conference is a professional gathering held annually at a pleasant location for all "SASaholics."* SAS users with diverse expertise and varying degrees of sophistication gather once a year at this conference

to exchange ideas and view demonstrations of the latest products developed by the SAS Institute. If you cannot attend this annual event, there are many regional or local SAS users groups to be involved with. Tips on connecting to a users group at the local or international level are found in Appendix A.

Seventh, the *SAS Institute will not forsake you when you are in trouble*. There are many helpful phone numbers, fax numbers, and subsidiaries or distributors located around the world to enable you to overcome obstacles. This information is found in Appendix A.

Last, but not the least, is that SAS is not as hard as you think. It is straightforward to enter data and set up data files in SAS. If you can type, you can learn SAS. So why not start the learning process now with this book?

1.2 What Is the History of the SAS Institute and Its Current Clientele?

SAS began in 1966 with the work of a computer scientist, A. J. Barr, at North Carolina State University. A year later, a colleague of his, Dr. J. H. Goodnight, contributed the regression analysis program to SAS. Thus, the basis for initial SAS releases was the general linear model. Since then, SAS has had eight major releases with additional products added to each release. In August of 2002, SAS acquired the powerful file transfer family of products (DBMS/COPY) from Conceptual Software Inc. (retrieved April 20, 2007, from www.conceptual.com), thus further strengthening its position as the leading software for general data analysis and information delivery. For detailed information about the company and its products, visit the company's Web site at www.sas.com.

Since its incorporation in July of 1976, the company has grown tremendously, with an unbroken trend of revenue increases each year of the Institute's history. The number of clients around the globe is more than 3.5 million. SAS products have been installed at more than 43,000 sites in more than 110 countries, including the dismantled U.S.S.R. and Eastern European regions. Approximately 75% of SAS installations have been for businesses including more than 90% of the 2007 *Fortune* 100 companies, 5% have been for educational institutions, and 20% for government agencies.

1.3 What Is the SAS System and What Can It Do? A Quick Walk-Through of SAS Products

SAS software is best described as an information delivery system suitable for an entire organization. As software, SAS represents a modular, integrated, and hardware-independent computing package. As an information

delivery system, SAS is capable of building a well-rounded, self-sufficient environment that is based on an organization's databases. Any trained data analyst can transform these data sets into useful information that is subsequently delivered to decision makers at the right moment to maximize the utility of the information. All these steps can be accomplished within SAS!

To attain the goal of efficient and timely information delivery, SAS is designed to (a) provide universal data access, (b) maximize the capabilities and breadth of applications in the software, (c) provide user interfaces, and (d) support and exploit the advantages of computer hardware platforms.

In simple words, SAS is a high-level programming language with flexible input/output styles and file management capabilities. It includes a wide range of analysis procedures to help users navigate through data so that the most succinct features in data may be readily transparent and subsequently analyzed. When faced with the challenge of dealing with multiple data formats or computing environments, you will find SAS capable of meeting such a challenge. This book, written with ample examples drawn from the social and behavioral sciences, will help you appreciate and master many useful features of SAS and their applications.

Behind the power of SAS is an array of versatile SAS products that are available for social and behavioral scientists. These products are called modules. SAS modules are capable of spreadsheet analysis, graphics, descriptive and inferential statistical analysis, report generation, database access, decision support, applications development, project management, client/server computing, executive information systems (i.e., EIS), and much more. These modules are sold separately or in a set. As far as computing environments are concerned, SAS integrates well with mainframes as well as with personal computers, including Macintoshes. Furthermore, a variety of UNIX operating systems are compatible with SAS. The following is a list of SAS products, commonly used by researchers and data analysts, along with their purposes and applications:

Purpose	Module	Contents
Connectivity to the SAS System	SAS/ASSIST	A menu-driven interface to the SAS System
Basic data processing and research tools	SAS/BASE	This is the core of the SAS System; it performs data management, analysis, and report writing
	SAS/FSP	Tool software for full-screen data entry, editing, querying, and letter writing

Purpose	*Module*	*Contents*
Statistical analysis	SAS/STAT	A collection of descriptive and inferential statistical procedures
	SAS/QC	Software for statistical quality improvement
	SAS/OR	Software for project management, operations research, scheduling, and linear programming
	SAS/LAB	A fully menu-driven module of guided data analysis in an experimental study
	SAS/ETS	Module for econometrics and time series analysis
	SAS/PH-Clinical	Software for reviewing and assimilating clinical trials data
Graphical representation of data or information	SAS/GRAPH	Software for information and representation of color graphics
	SAS/INSIGHT	Software for graphical data analysis
User-defined calculations	SAS/IML	An interactive matrix language for advanced numerical analysis
	SAS/CALC	Electronic spreadsheets
Linking the SAS system with other hardware environments, data-bases, or software	SAS/SHARE	A multiuser access tool for concurrent updating of SAS data libraries and data sets on mainframe systems
	SAS/ACCESS	Interfaces for linking the SAS System with various databases, such as ORACLE, CA-DATACOM, CA-IDMS, INGRES, INFORMATION, ADABAS, SYSTEM 2000 Data Management software, OS/2 Database Manager, etc.
	SAS/CONNECT	Facility for cooperative processing
Holistic systemwide application of SAS	SAS/EIS	Facility for building and maintaining executive information systems
	SAS/AF	An interactive facility with menus and screens for applications development, computer-based training, and online help systems
	SAS/TOOLKIT	Facility for extending the SAS System's capabilities
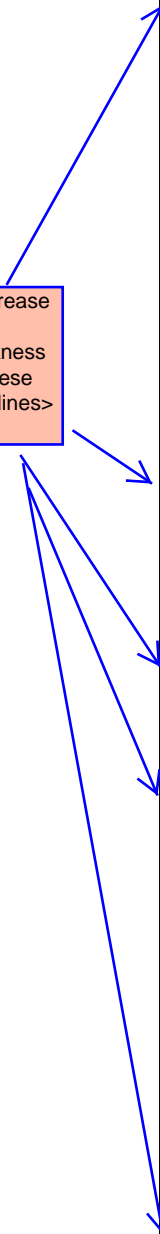
1.4 What Is the Scope and Style of This Book?

Topics covered in this book range from "Why should you learn SAS?" and "Where do you start?" to descriptive and inferential statistical analyses such as analysis of variance and regression. These topics are included to prepare you to deal with data frequently encountered in school projects, psychological experiments, social surveys, public polling, health research, clinical trials, marketing research, environmental studies, and many other contexts.

Specifically, the book is divided into three parts. Part I introduces SAS and basic data processing techniques (Chapters 1 through 7). Part II presents statistical procedures covered by most introductory statistics or research methods courses (Chapters 8 through 17). Part III deals with advanced data and file management (Chapters 18 through 20).

Chapters 1 and 2 are "starter" chapters that lay the groundwork for starting to learn more about data analysis and SAS. Chapters 3 to 5 address input and output formatting issues in data analysis. These issues include how you prepare your data for SAS to process, reading data stored in different formats by SAS, reading missing data in SAS, and displaying analysis results in SAS. Chapters 6 and 7 cover details of data processing such as recoding or transforming data, documenting SAS programs and output, registering variable types, verifying data, and debugging SAS programs.

Chapters 8 to 17 present SAS procedures (or commands in SPSS language) for descriptive and inferential statistical analyses. These procedures are for summarizing data, visually displaying data, analyzing categorical or continuous data, conducting average comparisons, and making predictions for individual observations. Conceptual explanations of statistical terms or equations are offered throughout these chapters, yet rigorous mathematical theories or proofs will not be presented.

Questions such as how to selectively and repetitively process data/variables/observations and how to manage data sets are dealt with in Chapters 18 to 20 (Part III).

Following these 20 chapters are three appendixes (A, B, and C). Appendix A contains information on reference books, hotlines, and a wealth of Internet resources that will help you continue to develop analytical and research skills with the power of SAS. Data sets used in this book are described in Appendix B; they are available from the SAGE Web site at www.sagepub .com/pengstudy, along with solutions (SAS programs and output) to all exercise questions at the end of each chapter. You may further develop your analytical competence by exploring different aspects of these data or raising different questions concerning the same data. Appendix C presents information on reading SPSS, Stata, Excel, Minitab, and SYSTAT data set files in the SAS System.

The writing style of this book is informal and conversational. Each data analysis technique is illustrated with at least one example that in turn can serve as a template for your own data analysis. The explanation of SAS

output is succinct and contextualized. Exercises provided at the end of each chapter will further develop your ability to reason with statistics and quantitative information. Solutions to these exercises are posted at the above Web site.

1.5 Summary

In this chapter, seven reasons were given for learning and mastering SAS. I hope that these reasons were convincing! Furthermore, a brief history of the SAS Institute (the brain behind the software!) and the dazzling array of its products were introduced. The products offer everything from data entry, data analysis, data reporting, data warehousing, and database management to natural language processing, specially designed to fully utilize the strengths of SAS.

This chapter ends with a description of the coverage of topics and the style of presentation adopted for this book. I hope that this chapter has aroused your interest in learning SAS for data analysis. Tricks published in this book will not only "convert" you to SAS but also hone your skills in making informed decisions and interpretations based on empirical data. Remember that SAS and computers have infinite patience. If you don't give up along the way, the mastery is ultimately yours!

So go for it!

1.6 Exercises

1. Look for additional resources at www.sagepub.com/pengstudy and bookmark this site.

2. Download the 10 data sets that accompany this book from www.sagepub.com/pengstudy and save these data sets on your computer hard drive or on a flash drive so that you may use them later when you begin to follow examples in this book.

3. Bookmark the SAS company site at www.sas.com and explore its various departments, especially the following:

   SAS Knowledge Base at http://support.sas.com/resources
   Technical Support at http://support.sas.com/techsup
   Training at http://support.sas.com/training/index.html
   SAS Discussion Forums at http://support.sas.com/forums/index.jspa
   Overall help at http://support.sas.com

4. Check out a local or the international SAS users group at http://support.sas.com/usergroups/intro.html

   Bookmark this site for future references.

5. Bookmark the online documentation for SAS version 9.1.3 at http://support.sas.com/ documentation/onlinedoc/91pdf/index_913.html

   The two documents that are particularly relevant to this book are

   SAS/STAT 9.1 User's Guide (SAS Institute Inc., 2004d)

   Base SAS 9.1.3 Procedures Guide, 2nd ed. (SAS Institute Inc., 2006a)

6. Join the SAS-L Listserv by following the instructions given in Appendix A of this book.

7. Use an Internet search engine to search for sites that advertise jobs for SAS data analysts in your discipline.