

Part 1 **The Basics**

-
- 1 Introduction**
 - 2 Getting Started**
 - 3 Creating SAS® Data Sets**
 - 4 Summarizing Data**

Chapter 1 **Introduction**

The *SAS System for Elementary Statistical Analysis* is designed to help you analyze your data using SAS software. This chapter covers

- the purpose of this book
- the audience for this book
- what this book is and isn't
- how this book is organized
- how to use this book
- what parts of SAS software you need.

Contents

<u>5</u>	<u>Purpose</u>
<u>5</u>	<u>Audience</u>
<u>5</u>	<u>What This Book Is and Isn't</u>
<u>6</u>	<u>How This Book Is Organized</u>
7	Using Part 1
8	Using Part 2
8	Using Part 3
9	Using Part 4
10	Using Part 5
10	Appendices, Summaries, and Other Helpful Information
11	Typographical Conventions
<u>11</u>	<u>How to Use This Book</u>
<u>12</u>	<u>What SAS Software Do You Need?</u>

Purpose

This book shows how to use SAS software to perform basic statistical analysis and explains the statistical methods used—when to use them, what the assumptions are, and how to interpret the results.

Audience

This book is for people who have minimal knowledge of SAS software or statistics or both. You may have extensive knowledge in your field as a manager, researcher, or analyst in industry. Or, you may be a student or professor in academia. The key is that you intend to use SAS software to analyze your data.

What This Book Is and Isn't

This book is a task-oriented tool that covers several basic types of data analysis. It integrates analysis methods with the SAS statements needed to perform the analysis and then explains the output in detail. Through the use of examples, it describes how the analysis leads to decisions about the data.

For those of you who know some statistical terms, the book covers descriptive summaries and graphs, two-way *t*-tests, analysis of variance, basic regression, and basic contingency table analyses. Examples of data analyses not discussed in this book are analysis of variance for designs other than a completely randomized design, multivariate analyses, time series methods, quality control, and operations research analyses.

This book shows how to use SAS software to perform the data analysis methods discussed. Because SAS software is a versatile and powerful language, often, an analysis can be performed another way with the software. This book does not cover all the ways SAS software can do a particular analysis.

This book concentrates on data analysis. However, some statistical background is needed in order for the discussions on analysis to make sense. Although you can perform analyses using only the information in this book, you may also want to refer to statistical texts for more background theory or detailed information about the methods. In addition, the book includes very few formulas. Formulas for the statistical

methods that are included in this book are available in many texts, as well as in other documentation for SAS software.

This book does not include methods for designing a study. Several references for the practical aspects of study design are given in the “Further Reading” sections at the end of Chapters 3-11. Designing your study is as important as choosing the correct analysis. No amount of high-powered statistical analysis will “fix” a study by answering questions that the study wasn’t designed to answer. If you are uncertain about how to design your study, you should consult a statistician or one of the references given in “Further Reading” before you begin. You should also think ahead about how you record your data. You can save yourself a lot of time by planning to record the data in a way that can be easily read into SAS.

This book is designed to:

- be task-oriented
- explain basic steps to create a SAS data set
- discuss several SAS tools for statistical analyses
- discuss SAS plots and graphs as appropriate.

This book is not designed to:

- replace SAS manuals or statistics texts
- explain all the ways to do a task in SAS
- discuss experimental design
- discuss advanced statistical analyses.

How This Book Is Organized

This book is divided into five parts.

Part 1: The Basics

- Chapter 1 Introduction
- Chapter 2 Getting Started
- Chapter 3 Creating SAS Data Sets
- Chapter 4 Summarizing Data

Part 2: Statistical Background

- Chapter 5 Understanding Some Basic Statistical Concepts
- Chapter 6 Estimating the Mean

Part 3: Comparing Groups

- Chapter 7 Comparing Two Groups
- Chapter 8 Comparing More Than Two Groups

Part 4: Fitting Lines to Data

- Chapter 9 Correlation and Regression
- Chapter 10 Basic Regression Diagnostics

Part 5: Data in Tables

- Chapter 11 Creating and Analyzing Crosstabulations.

The first part of the book shows you how to get your data into the computer. The second part of the book gives you a statistical foundation for the analyses discussed in Parts 3, 4, and 5. These other parts of the book show you how to perform various analyses. Chapters are as independent as possible—you don't need to absorb the contents of several chapters to perform only one type of analysis.

Using Part 1

Part 1 is essential for all readers who don't have experience with SAS software. Chapter 1 contains a brief introduction. Chapter 2 tells you what you need to know before you can use SAS software and gives an introduction to the SAS System. Chapter 3 describes how to create a SAS data set. Chapter 4 shows how to use SAS software to get descriptive statistics, frequency tables, bar charts, and some simple plots of the data. The output is explained in detail.

Starting with Chapter 2, there are three summaries and a reading list at the end of each chapter. The "Key Ideas" summary contains the main ideas of the chapter. The "Syntax" summary gives the general form for all SAS statements discussed in the chapter. (There is one exception: the SAS statements to create a SAS data set are summarized in Chapter 3 only.) The "Example" summary lists the SAS statements used to produce all the output that is shown in the chapter. The reading list contains a list of books that give more information about the topics discussed in the chapter.

Starting with Chapter 4, this second edition of the book shows how to summarize or analyze the data with SAS/INSIGHT, as appropriate. This

interactive graphical tool is used more and more for data investigation. However, since it requires more than just the basic SAS package, we show the SAS/INSIGHT methods as an addition to other methods.

Using Part 2

Part 2 focuses on statistical concepts and provides a foundation for analyses presented in Parts 3, 4, and 5. Some of the ideas discussed include populations and samples, the normal distribution, the Empirical Rule, hypothesis testing, the effect of sample size and population variance on estimates of the mean, the Central Limit theorem, and confidence intervals for the mean. For learning SAS to perform statistical tasks, the chapters discuss the following:

Chapter	Statistical Tasks in SAS
5	Testing for normality
6	Calculating confidence intervals

Using Part 3

Part 3 shows how to compare groups of data. The list below gives examples of the types of problems you can analyze by using the methods presented in Part 3.

- Your new employees must take a proficiency test before they may operate a dangerous and expensive piece of equipment. Some of your new employees are experienced with this equipment and some are not. You want to find out if the average test scores for the two groups (experienced and inexperienced) are different.
- You measured the heart rate of students before and after mild exercise on a treadmill. You think the “before” heart rates are likely to be lower than the “after” heart rates. You want to know if the changes in heart rates are greater than could happen by chance.
- You conducted a study to compare the effects of five fertilizers on potted geraniums. At the end of six weeks, you measured the height of the geranium plants. Now, you want to compare the average heights for the plants that received the five different fertilizers.
- You recorded the number of defective items from an assembly line on the workdays Monday through Friday. You want to determine if there are differences in the number of defective items made on different days of the week.

For learning SAS to perform statistical tasks, the chapters discuss the following:

Chapter Statistical Tasks in SAS

- | | |
|---|--|
| 7 | two-sample <i>t</i> -test
paired difference <i>t</i> -test
Wilcoxon Signed-rank test
Wilcoxon Rank-sum test |
| 8 | Analysis of variance (<i>ANOVA</i>)
Kruskal-Wallis test
Multiple comparison procedures |

In addition, both chapters discuss methods for summarizing and graphing data from two or more groups.

Using Part 4

Part 4 shows how to fit a line or curve to a set of data. Here are examples of problems you can analyze when using methods from Part 4.

- You want to determine the relationship between weight and age in children. You want to fit a line or curve to define the relationship. You also want to produce a plot showing the data and the fitted line or curve. Finally, you want to examine the data to find atypical points that are quite far from the fitted line.
- You have varied the temperature of an oven in a manufacturing process. You want to determine the amount of change in hardness of the product that is caused by a given amount of change in temperature. You also want to be able to predict hardness of future products made using a given temperature.

For learning SAS to perform statistical tasks, the chapters discuss the following:

Chapter Statistical Tasks in SAS

- | | |
|----|--|
| 9 | Correlations
Straight-line regression
Multiple regression
Prediction and confidence limits |
| 10 | Plotting residuals and predicted values
Plotting residuals and independent variables
Plotting residuals in time sequence |

In addition, the chapters discuss investigation of outliers in a regression analysis.

Using Part 5

Part 5 consists of Chapter 11, which shows some basic methods for analyzing data in tables. The list below gives examples of problems you can solve using methods from this chapter.

- You have collected data on the colors of shampoo preferred by men and women. You want to know if the way color preferences are distributed is the same for men and women.
- You have conducted an opinion poll to evaluate how people in your town feel about the proposal for a new tax to build a sports arena. You want to find out if the people who live close to the site for the proposed arena feel differently about the arena than do the people who live farther away.
- Your company is conducting a test-market for a new product. You have conducted a survey in which you asked shoppers if they had bought the product. You also asked a series of questions designed to get opinions about the packaging of the product. You want to find out if these opinions are different between people who bought the new product and people who have not.

For learning SAS to perform statistical tasks, the chapter discusses the following:

Chapter	Statistical Tasks in SAS
11	Chi-square test for independence Measuring correlation

In addition, the chapter explains how to summarize data in tables.

Appendices, Summaries, and Other Helpful Information

Other helpful features are the appendices and the summaries. Appendix 1, “Display Manager Basics,” gives a brief introduction to the SAS Display Manager. Appendix 2, “Troubleshooting,” provides solutions to common problems in SAS programs. Appendix 3, “Additional Information,” tells how to get publications and training catalogs from SAS Institute and describes some of the additional information and services available. Appendix 4, “Syntax Summary,” contains the syntax summaries from the ends of Chapters 2-11. The “Quick Reference Table” on the inside front cover provides a place for you to write down some specifics about the computer system you use. A glossary and an index are at the end of the book.

Typographical Conventions

The typographic conventions that are used in this book are

roman	used for most text.
<i>italics</i>	used to define new terms, indicate items in SAS statements that you need to supply, and point to items in the output.
bold	used in the general form of SAS statements, bold type indicates that you must use the exact spelling and punctuation shown . In addition, headings and phrases and sentences of extreme importance are in bold type.
code	used to show examples of SAS statements.

How to Use This Book

If you have never used SAS software before, first read all of Part 1. These chapters present the basics of using the software to create and summarize a SAS data set. Then, if you aren't familiar with statistics, read Part 2. Otherwise, you can skip to other parts of the book. The descriptions in "How This Book Is Organized" should give you a good idea of the chapters that best fit your problem.

If you have used SAS software before but are not familiar with statistics, skim Part 1. Then read Part 2, which gives an explanation of several statistical concepts. Then skip to other parts of the book that best meet your needs.

If you have used SAS software before and are familiar with statistics but don't know how to handle your particular problem, first skim Part 1. Then go to the chapter that best meets your needs to learn how to use the software for your problem.

Do not use this book as a substitute for a statistical text. While the book shows how to use SAS software to perform certain analyses and describes the statistical methods, the SAS statements, and the SAS output, it does not attempt to replace statistical texts.

What SAS Software Do You Need?

To do most of the summaries and analyses in the book, you need base SAS software and SAS/STAT software. To do the graphs shown in the book, you need SAS/GRAPH software. If you don't have SAS/GRAPH, you can get low-resolution graphs using base SAS software.

Several chapters show how to perform tasks using base SAS and SAS/STAT and then include an optional PROC INSIGHT section. For these summaries and analyses, you need SAS/INSIGHT software.