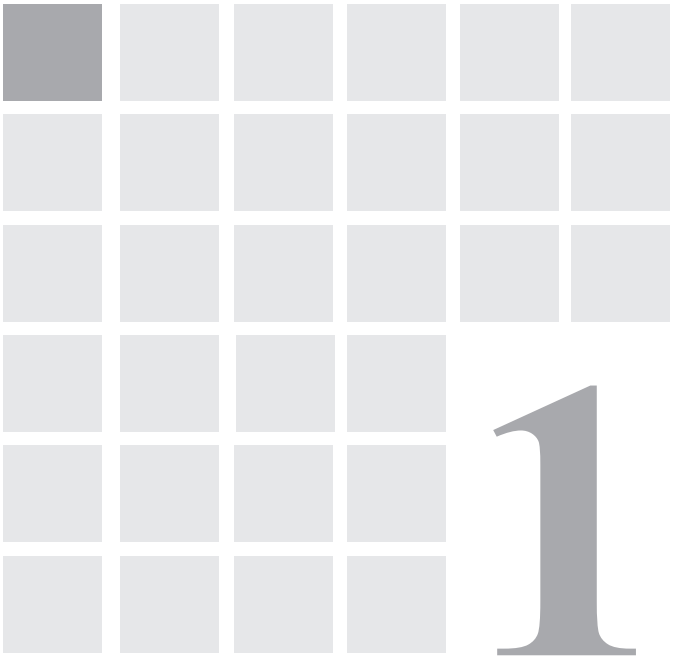

Using This *Student Guide*

Introduction

Overview

This chapter introduces you to the SAS System, a computer application that can be used to perform statistical analyses. It explains just what SAS is, where it is installed, and describes some of the advantages associated with using SAS for data analysis. Finally, it briefly summarizes what you will learn in each of the chapters that comprise this *Student Guide*.

Intended Audience and Level of Proficiency

This guide is intended for those who want to learn how to use SAS to perform elementary statistical analyses. The guide assumes that many students using it have not already taken a course on elementary statistics. To assist these students, this guide briefly reviews basic terms and concepts in statistics at an elementary level. It was designed to be easily understood by first and second year college students.

This book was also designed to be user-friendly to those who may have little or no experience with personal computers. The beginning of Chapter 3, "Tutorial: Using the SAS Windowing Environment to Write and Submit SAS Programs," reviews basic concepts in using Microsoft Windows, such as selecting menus, double-clicking icons, and so forth. Those who already have experience in using Windows will be able to quickly skim through this elementary material.

Platform and Version

This guide shows how to use the SAS System for Windows, as opposed to other operating environments. This is most apparent in Chapter 3, "Using the SAS Windowing Environment to Write and Submit SAS Programs." However, the remaining chapters show how to write SAS *code* to perform statistical analyses, and most of this material will be useful to all SAS users, regardless of the operating environment. This is because, for the most part, the same SAS code can be used on a wide variety of operating environments to obtain the same results.

This book was designed for those using the SAS System Version 8 and later versions. It may also be helpful to those using earlier versions of SAS (such as V6 or V7). However, if you are using one of these earlier versions, it is likely that some of the SAS system options described here are not available with your version. It is also likely that some of the SAS output that you obtain will be arranged differently than the output that is presented here.

Materials Needed

To complete the activities described in this book, you will need

- access to a personal computer on which the SAS System for Windows has been installed,
- one (and preferably two) 3.5-inch disks, formatted for IBM PCs (or some other type of storage media).

Some students using this book will also use its companion volume, *Step-by-Step Basic Statistics Using SAS: Exercises*. The chapters in the *Exercises* book parallel most of the chapters contained in this *Student Guide*. Each chapter in the *Exercises* book contains two assignments for students to complete. Complete solutions are provided for the odd-numbered exercises, but not for the even-numbered ones. The *Exercises* book can give you useful practice in learning how to use SAS, but it is not absolutely required.

Introduction to the SAS System

Why Do You Need This *Student Guide*?

This *Student Guide* shows you how to use a computer application called the SAS System to perform elementary statistical analyses. Until recently, students in elementary statistics courses typically performed statistical computations by hand or with a pocket calculator. In recent years, however, the increased availability of computers has made it possible for students to also use statistical software packages such as SPSS and the SAS System to perform these analyses. This latter approach allows students to focus more on conceptual issues in statistics, and spend less time on the mechanics of performing mathematical operations by hand. Step by step, this *Student Guide* will introduce you to the SAS System, and will show you how to use it to perform a variety of statistical analyses that are commonly used in the social and behavioral sciences and in education.

What Is the SAS System?

The SAS System is a modular, integrated, and hardware-independent application. It is used as an information delivery system by business organizations, governments, and universities worldwide.

SAS is used for virtually every aspect of information management in organizations, including decision support, project management, financial analysis, quality improvement, data warehousing, report writing, and presentations. However, this guide will focus on just one aspect of SAS: its ability to perform the types of statistical analyses that are appropriate for research in the social sciences and education.

By the time you have completed this text, you will have accomplished two objectives: you will have learned how to perform elementary statistical analyses using SAS, and you will have become familiar with a widely used information delivery system.

Who Uses SAS?

The SAS System is widely used in business organizations and universities. Consider the following statistics from July 2002:

- SAS supports over 40 operating environments, including Windows, OS/2, and UNIX.
- SAS Institute's computer software products are installed at over 38,400 sites in 115 countries.
- Approximately 71% of SAS installations are in business locations, 18% are education sites, and 11% are government sites. It is used for teaching and research at about 3,000 university locations.
- It is estimated that SAS software products are used by more than 3.5 million people worldwide.
- 90% of all Fortune 500 companies are SAS clients.

Using the SAS System for Statistical Analyses

SAS is a particularly powerful tool for social scientists and educators because it allows them to easily perform virtually any type of statistical analysis that may be required in their research. SAS is comprehensive enough to perform the most sophisticated multivariate analyses, but is so easy to use that undergraduates can perform simple analyses after only a short period of instruction.

In a sense, the SAS System may be viewed as a library of prewritten statistical algorithms. By submitting a brief SAS program, you can access a procedure from the library

and use it to analyze a set of data. For example, below are the SAS statements used to call up the algorithm that calculates Pearson correlation coefficients:

```
PROC CORR   DATA=D1;
   RUN;
```

The preceding statements will cause SAS to compute the Pearson correlation between every possible pair of numeric variables in your data set. Being able to call up complex procedures with such a simple statement is what makes SAS so powerful. By contrast, if you had to prepare your own programs to compute Pearson correlations by using a programming language such as FORTRAN or BASIC, it would require many statements, and there would be many opportunities for error. By using SAS instead, most of the work has already been completed, and you are able to focus on the results of the analysis rather than on the mechanics of obtaining those results.

Contents of This *Student Guide*

Overview

This guide has two objectives: to teach the basics of using SAS in general and, more specifically, to show how to use SAS procedures to perform elementary statistical analyses. Chapters 1–4 provide an overview to the basics of using SAS. The remaining chapters cover statistical concepts in a sequence that is representative of the sequence followed in most elementary statistics textbooks.

Chapters 10–17 introduce you to inferential statistical procedures (the type of procedures that are most often used to analyze data from research). Each chapter shows you how to conduct the analysis from beginning to end. Each chapter also provides an example of how the analysis might be summarized for publication in an academic journal in the social sciences or education. For the most part, these summaries are written according to the guidelines provided in the *Publication Manual of the American Psychological Association* (1994).

Many students using this book will also use its companion volume, *Step-by-Step Basic Statistics Using SAS: Exercises*. For Chapters 3–17 in this student guide, the corresponding chapter in the exercise book provides you with a hands-on exercise that enables you to practice the data analysis skills that you are learning.

The following sections provide a summary of the contents of the remaining chapters in this guide.

Chapter 2: Terms and Concepts Used in This Guide

Chapter 2 defines some important terms related to research and statistics that will be used throughout this guide. It also introduces you to the three types of files that you will work with during a typical session with SAS: the SAS program, the SAS log, and the SAS output file.

Chapter 3: Tutorial: Using the SAS Windowing Environment to Write and Submit SAS Programs

The **SAS windowing environment** is a powerful application that you will use to create, edit, and submit SAS programs. You will also use it to review your SAS logs and output. Chapter 3 provides a tutorial that teaches you how to use this application. Step by step, it shows you how to write simple SAS programs and interpret their results. By the end of this chapter, you should be ready to use the SAS windowing environment to write and submit SAS programs on your own.

Chapter 4: Data Input

Chapter 4 shows you how to use the DATA and INPUT statements to create SAS data sets. You will learn how to read both numeric and character variables by using a simple, list style for data input. By the end of the chapter, you will be prepared to input the data sets that will be presented throughout the remainder of this guide.

Chapter 5: Creating Frequency Tables

Chapter 5 shows you how to create frequency tables that are useful for understanding your data and answering some types of research questions. For example, imagine that you ask a sample of 150 people to tell you their age. If you then used SAS to create a frequency table for this age variable, you would be able to easily answer questions such as

- How many people are age 30?
- How many people are age 30 or younger?
- What percent of people are age 45?
- What percent of people are age 45 or younger?

Chapter 6: Creating Graphs

Chapter 6 shows you how to use SAS to create frequency bar charts—bar charts that indicate the number of people who displayed a given value on a variable. For example, imagine that you asked 150 people to indicate their political party. If you used SAS to create a frequency bar chart, the resulting chart would indicate the number of people who are democrats, the number who are republicans, and the number who are independents.

Chapter 6 also shows how to create bar charts that plot subgroup means. For example, assume that, in the "political party" study described above, you asked the 150 subjects to indicate both their political party and their age. You could then use SAS to create a bar chart that plots the mean age for people in each party. For instance, the resulting bar chart might show that the average age for democrats was 32.12, the average age for republicans was 41.56, and the average age for independents was 37.33.

Chapter 7: Measures of Central Tendency and Variability

Chapter 7 shows you how to compute measures of variability (e.g., the interquartile range, standard deviation, and variance) as well as measures of central tendency (e.g., the mean, median, and mode) for numeric variables. It also shows how to use stem-and-leaf plots to determine whether a distribution is skewed or approximately normal in shape.

Chapter 8: Creating and Modifying Variables and Data Sets

Chapter 8 shows how to use subsetting IF statements to create new data sets that contain a specified subgroup from the original sample. It also shows how to use mathematical operators and IF-THEN statements to recode variables and to create new variables from existing variables.

Chapter 9: Standardized Scores (*z* Scores)

Chapter 9 shows how to transform raw scores into standardized variables (*z* score variables) with a mean of 0 and a standard deviation of 1. You will learn how to do this by using the data manipulation statements that you learned about in Chapter 8. Chapter 9 also illustrates how you can review the sign and absolute magnitude of a *z* score to understand where a particular observation stands on the variable in question.

Chapter 10: Bivariate Correlation

Bivariate correlation coefficients allow you to determine the nature of the relationship between two numeric variables. Chapter 10 shows you how to use the CORR procedure to compute Pearson correlation coefficients for interval- and ratio-level variables. You will also learn to interpret the *p* values (probability values) that are produced by PROC CORR to determine whether a given correlation coefficient is significantly different from zero. Chapter 10 also shows how to use PROC PLOT to create a two-dimensional scattergram that illustrates the relationship between two variables.

Chapter 11: Bivariate Regression

Bivariate regression is used when you want to predict scores on an interval- or ratio-level criterion variable from an interval- or ratio-level predictor variable. Chapter 11 shows you how to use the REG procedure to compute the slope and intercept for the regression equation, along with predicted values and residuals of prediction.

Chapter 12: Single-Sample *t* Test

Chapter 12 shows how to use the TTEST procedure to perform a single-sample *t* test. This is an inferential procedure that is useful for determining whether a sample mean is significantly different from a specified population mean. You will learn how to interpret the *t* statistic, and the *p* value associated with that *t* statistic.

Chapter 13: Independent-Samples *t* Test

You use an independent-samples *t* test to determine whether there is a significant difference between two groups of subjects with respect to their mean scores on the dependent variable. Chapter 13 explains when to use the equal-variance *t* statistic versus the unequal-variance *t* statistic, and shows how to use the TTEST procedure to conduct this analysis.

Chapter 14: Paired-Samples *t* Test

The paired-samples *t* test is also appropriate when you want to determine whether there is a significant difference between two sample means. The paired-samples approach is indicated when each score in one sample is dependent upon a corresponding score in the second sample. This will be the case in studies in which the same subjects provide repeated measures on the same dependent variable under different conditions, or when matching procedures are used. Chapter 14 shows how to perform this analysis using the TTEST procedure.

Chapter 15: One-Way ANOVA with One Between-Subjects Factor

One-way analysis of variance (ANOVA) is an inferential procedure similar to the independent-samples *t* test, with one important difference: while the *t* test allows you to test the significance of the difference between two sample means, a one-way ANOVA allows you to test the significance of the difference between more than two sample means. Chapter 15 shows how to use the GLM procedure to perform a one-way ANOVA, and then to follow with multiple comparison (post hoc) tests.

Chapter 16: Factorial ANOVA with Two Between-Subjects Factors

A one-way ANOVA, as described in Chapter 15, may be appropriate for analyzing data from an experiment in which the researcher manipulates only one independent variable. In contrast, a factorial ANOVA with two between-subjects factors may be appropriate for analyzing data from an experiment in which the researcher manipulates two independent variables simultaneously. Chapter 16 shows how to perform this type of analysis. It provides examples of results in which the main effects are significant, as well as results in which the interaction is significant.

Chapter 17: Chi-Square Test of Independence

Nonparametric statistical procedures are procedures that do not require stringent assumptions about the nature of the populations under study. Chapter 17 illustrates one of the most common nonparametric procedures: the chi-square test of independence. This test is appropriate when you want to study the relationship between two variables that assume a limited number of values. Chapter 17 shows how to conduct the test of significance and interpret the results presented in the two-way classification table created by the FREQ procedure.

References

Many statistical procedures are illustrated in this guide by showing you how to analyze fictitious data from an empirical study. Many of these "studies" are loosely based on actual investigations reported in the research literature. These studies were chosen to help introduce you to the types of empirical investigations that are often conducted in the social and behavioral sciences and in education. The "References" section at the end of this guide provides complete references for the actual studies that inspired the fictitious studies reported here.

Conclusion

This guide assumes that some of the students using it have not yet completed a course on elementary statistics. This means that some readers will be unfamiliar with terms used in data analysis, such as “observations,” “null hypothesis,” “dichotomous variables,” and so on. To remedy this, the following chapter, "Terms and Concepts Used in This Guide," provides a brief primer on basic terms and concepts in statistics. This chapter should lay a foundation that will make it easier to understand the chapters to follow.