Chapter 1

Touring the Wonderful World of SAS

*O*ne of the questions newcomers ask most frequently about SAS is "What does the name mean?" After all, those capital letters usually indicate an acronym, right? Today, SAS just refers to the name of a company. If you've been around the world of data analysis for a while, however, you may also be familiar with the old meaning of the abbreviation, *Statistical Analysis System*.

SAS software was initially developed by a bunch of really smart and inquisitive people at North Carolina State University (NCSU) in the late 1960s and early 1970s. Some of these people are still at SAS (the company) as owners or executives: Jim Goodnight (the current company president), John Sall, and Herb Kirk (the first SAS user). Most of these SAS software pioneers were trained as statisticians or mathematicians and developed the SAS language to specifically help analyze a variety of scientific experiments being conducted at NCSU and other research universities.

Over time, the software became as important as the experiments it was being used to analyze. In 1976, several people brave enough to leave the cozy life of academia for the then unknown world of software started the company today known as SAS. The first few years were a bit rough; but before long, word of this software and its capabilities began to spread, revenues increased, and the company began to grow.

This chapter is an overview of the power and flexibility of SAS for a wide range of applications and industries. In particular, SAS has expanded from being just for experts to meeting the needs of a wide range of users in almost every industry and country in the world. SAS has come a long way, changing from just a programming language to a wide range of applications tailored to various business and scientific needs.

SAS — Isn't That Just for Gurus?

You might assume that you need to be a statistician or math guru to use SAS, but that's happily not the case. In the last few years, SAS has made significant investment in taking the unparalleled analytical and data management capabilities developed over 30 years and making them available to almost anyone with a problem to solve in business, science, or government. With new products such as SAS Enterprise Guide and the SAS Add-In for Microsoft Office, SAS has never been more accessible or flexible. These products provide user-friendly interfaces and wizards to maximize the heavy-duty capabilities that SAS has long provided to gurus!
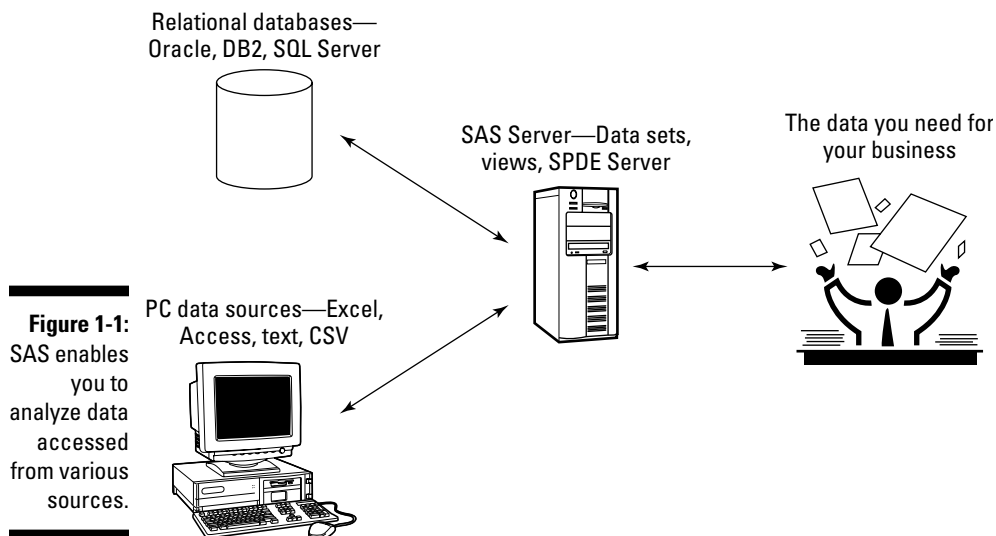
Most of this book is dedicated to simple-to-learn principles that are full of possibilities and limited only by your situation and imagination. SAS offers so much potential that this book just scratches the surface and hopefully gets you up to speed on the basics.

Data, Data Everywhere — But Not Where I Need It!

The glamorous side of business intelligence and data analysis is all the gee-whiz reports, graphs, and impressive statistics you can present. (It must be true because my p-value says so! And don't worry if you don't know what a *p-value* is right now. That will come later.) The surprising secret of actually arriving at good results for decision-making is the huge amount of time that many people spend accessing, organizing, and preparing their data for a particular analysis. In visiting more than 50 major companies in our various work experiences, we've found that the common theme is the massive amount of resource and rework time spent on the data preparation aspect of business analytics.

As mentioned earlier in this chapter, the first developers of SAS were doing real-world research projects and faced these very same data preparation and analysis issues. Consequently, they developed products that allow seamless

access to over 100 data sources on almost every computing platform currently in use. This capability was way ahead of its time back then and is still hands-down the best we have used. These data access products — SAS/ACCESS products — run on your SAS server. They allow fast, seamless access to disparate data sources for your analysis (see Figure 1-1).

Figure 1-1: SAS enables you to analyze data accessed from various sources.

Relational databases— Oracle, DB2, SQL Server

SAS Server—Data sets, views, SPDE Server

The data you need for your business

PC data sources—Excel, Access, text, CSV

One real-life data preparation story

At one prominent aerospace company, Six Sigma black-belts reported that 85 percent of their time is spent collecting, cleaning, and preparing their data for the business tasks at hand. Even worse, they realize this work is duplicated across various departments. They all end up doing the same preparation work with a given data source, such as data describing all products currently sold, their predecessor products, and the dates that products were discontinued.

This data resides on different platforms in various formats with a wide array of data rules. Staff work with older data in text files on a mainframe computer, data from an acquired subsidiary in Oracle on UNIX, data in DB2 from a new ERP system (Enterprise Resource Planning system — SAP in this case) on a Windows server, or data in a spreadsheet on someone's PC. When each team brings this data together for its own projects, they often arrive at different results. Upper management wonders why teams can never agree on even basic metrics and the analyses needed to run the business. . . .

SAS can get to the data, but that's only the beginning. SAS also has excellent tools to enable centralized management of your data. Applications such as SAS Enterprise Guide have a wide array of data access, query, and management capabilities that enable you to slay your data-management dragons in a flexible and effective manner.

SAS also offers the SAS Data Integration Server, which enables your power users to effectively access, manage, and aggregate your data. The SAS Data Integration Server focuses around the types of problems commonly connected to data warehousing and data quality. Using the SAS Data Integration Server allows you to have one integrated view of your data that is built on common rules and assumptions. The value here is in avoiding different answers to the same question by ensuring that everyone has access to a user-friendly, consistent data store. You find out more about this topic in Chapter 16.

Data Summaries and Reporting

If you've worked with traditional business intelligence tools from other software vendors, you might be familiar with data summarization and reporting. These tasks are critical to your ability to pull value from the data and knowledge inherent in your organization. Unfortunately, this immediate need for data is often the only area that people focus on when they ask for information to answer a particular question. If you can take a broader, long-term approach to your data management, reporting, and analysis needs, however, you can save money and time while yielding superior results.

One example to illustrate this point is a report of accounts past due. You could generate this information in Microsoft Excel and copy and paste subsets of the data to send to various sales teams. This is a very manual process. Or, you could design a report that can be easily updated with the latest data. This report can use subsets for accounts for each sales team and link to order details for each overdue account to show exactly what was in the overdue order. Also imagine if this report could be delivered automatically over the Web, by e-mail, or directly into Microsoft Office. Now it is a much more flexible and powerful asset — all available from one SAS report!

Why summarize data?

Elsewhere in this chapter we give an example of reducing a 50-million–row table to just a few rows. Imagine, then, that you want to present that data, summarized in three forms: a pie chart, a listing, and a bar chart. By explicitly summarizing the large data source once (collapsing the 50 million rows down to 100 or so rows) and then creating the pie chart, listing, and bar chart from the summarized form, you get a much quicker generation of your results for your analyses.

Some simple forms of data summaries include sums, averages, medians, ranges, counts (sometimes called *frequencies*), and percentages. If you're interested in determining total sales by region, for example, the data source you have with this information might be a 50-million–row table. By using the summary functions of SAS, you can collapse this data to a small number of rows — one row per region, for example. Many functions in SAS automatically summarize the data for you. A pie chart of the sales by region would also automatically collapse the data to just a few rows before charting it.

SAS has a variety of powerful techniques to summarize your data, from basic counts, means, medians, minimum values, and maximum values, to sophisticated algorithms that allow you not just to aggregate the data but actually find relevant confidence intervals around the aggregations you request.

The Secret Sauce: Analytics to Optimize the Present and Predict the Future

If you were familiar with SAS before you started reading this chapter, you may be aware that SAS was made famous by its analytic capabilities. And you may be wondering whether you can easily use the analytic capabilities that SAS offers. Even if they are easy to use, can they really make a difference in your business? We can almost absolutely, positively guarantee that the answer will be *Yes*! (Okay, legalese time. This is not some binding guarantee. Your results and mileage can vary, but we're 99.999% sure.)

Almost every analytic technique, statistic, and test is designed to help better identify the true state of something by analyzing limited information. Here are some examples of where analytics can come in handy:

✔ Did the Western sales region really have a better average sales nunber than the Eastern region?

✔ Do customers who buy our gum spend more money at retailers than customers who don't?

✔ What are the projected sales over the next year of CinnaPecans if I lower their price by 10 percent?

✔ Which customer demographic factors are useful in predicting customers' receptiveness to a direct marketing solicitation?

To answer any of these questions effectively, you first need access to data that is of high quality, familiar to you, and properly organized so that you can apply the appropriate analytic technique for the question at hand. Even after you apply the appropriate analytic technique, you need an integrated way to evaluate the success of the technique and a method of presenting (reporting) the results so that even managers (like us) can understand.

Table 1-1 offers a high level view of some of the analytic capabilities from SAS, their potential applications, and where you can go to in this book to find out more about the technique.

Table 1-1	Example Applications of SAS Analytics	
Real World Example	**Statistical Technique**	**Chapter**
An engineer wants to predict the mean time until failure for a new LED television based on 15 test models	Survival Analysis	Chapter 9
A manager wants to know the impact on projected sales next year if she doubles marketing spending	Forecasting	Chapter 9
A clinician wants to know the effect on patient response of doubling the dose of a new drug	Mixed Models	Chapter 8
A sales manager wants to know the projected profitability of a new customer based on the customer's demographic profile	Data Mining	Chapter 9
A taste tester wants to know if people really prefer Fizzy Cola over Foamy Cola	Categorical Data Analysis	Chapter 9

Real World Example	Statistical Technique	Chapter
A procurement team wants to test whether the new super strong titanium bolts meet the specified strength specs for its new jet	Quality Control	Chapter 9
A sales promotion manager at OmniLoMart and her team want to know projected sales by country, store, and even SKU for the next week	High Performance Forecasting	Chapter 9
A hospital wants to predict patient stay length based on physician and nurse comments captured in the patient database	Text Mining	Chapter 9

Sharing the SAS Wealth

SAS has gone above and beyond its traditional market in the last few years to add an impressive array of tools and delivery mechanisms to make the lives of business analysts, managers, and executives easier and more productive. The following list describes just a few of the tools SAS offers you:

✔ **OLAP (On-Line Analytic Processing):** Frequently referred to by lay people as *pivot tables*, provides a mechanism for large volumes of data to be summarized in advance and presented to users via customized tools specifically designed to make exploring this data easy and fast. With OLAP, you can take a very large table, such as a sales history table for a large retailer, and predefine certain categories and metrics of interest that are run on a nightly basis. This results in a greatly collapsed data size with data stored in a specific format that enables very fast creation of summaries and exploration. Figure 1-2 illustrates a view of such a sales table before and after using it in OLAP.

✔ **SAS Add-In for Microsoft Office:** Provides you with direct access to SAS reports, data engines, data management, reporting, and analytic tasks seamlessly from Microsoft Excel, Word, and PowerPoint. The add-in allows you to avoid the commonly discussed concern of using Excel for your analytic needs, often called *spreadsheet hell.* Spreadsheet hell is the issue of using a simplistic yet user-friendly tool like Excel for complex data processing that really should be done with a better tool. SAS is well suited to perform this type of processing through the SAS add-in. When you use the SAS add-in, SAS content and data sources are centrally maintained and can be dynamically synchronized with your SAS server to ensure that all analysts in your company are accessing "one version of

the truth." A simple example is illustrated in Figure 1-3: a centrally cre-
ated and maintained SAS forecast analysis that is dynamically streamed
and easily updated by end users from PowerPoint.

✔ **SAS Information Delivery Portal and SAS Web Report Studio:** Allow
for simplified delivery of content over the Web and intuitive reporting
for almost any level of user. Users access a centrally maintained view of
their data to quickly create powerful and insightful reports that can be
easily shared throughout the organization. Figure 1-4 illustrates just one
of the many report formats that you can create in a matter of minutes
with SAS Web Report Studio.

Figure 1-2:
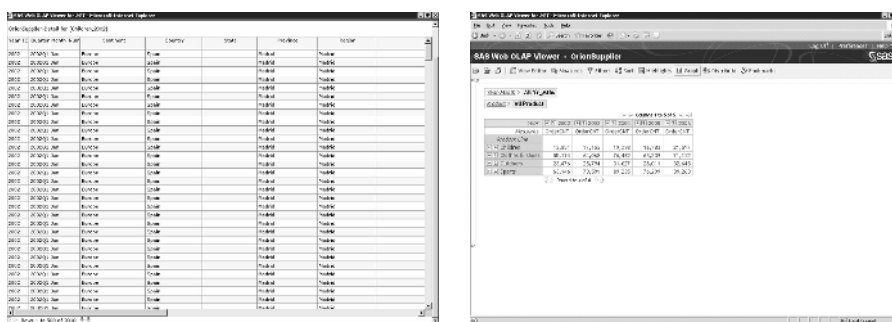Sales data
in standard
table form
and then in
OLAP form.

Figure 1-3:
Using
the SAS
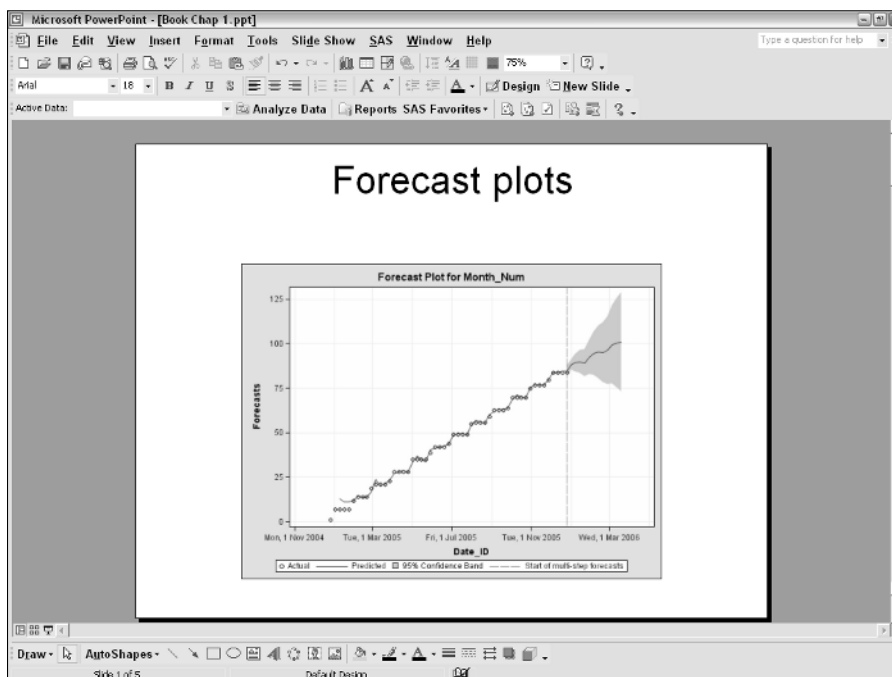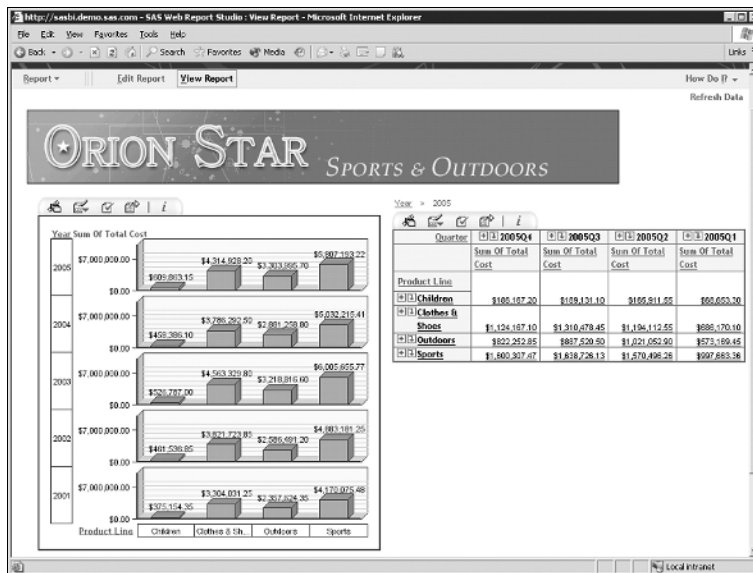add-in in
PowerPoint
for a SAS
Forecast
report.

Figure 1-4:
Using SAS
Web Report
Studio for a
Web-based
report.

What the IT Department Needs to Know

The ease of use and the powerful analytical applications of SAS are great for the end user and number crunchers, but SAS is friendly to your IT folks as well. The good news for IT professionals is that SAS 9 provides you with a centralized approach for deploying software, managing security, managing user environments, creating content, distributing content, and controlling user access to data.

By using standard software packaging tools, administrators can prepackage the distribution of SAS software. Using SAS Management Console to maintain metadata in the SAS Metadata Server (also known as the Open Metadata Repository), you can configure servers, server options, users, and user groups, manage data sources, and manage content available to users.

SAS Information Map Studio enables you to create dynamic data views based on administrator-defined Information Maps. These Information Maps hide the complexity and danger of accessing complex data schemas by presenting users with administrator-defined business views of the data. Based on user selections, SQL is dynamically created to provide them with just the data they need for their report.

More details that IT folks may be interested in reviewing are covered in Chapters 4 (data access), 15 (setting up SAS), 16 (data warehousing), 17 (SAS programming with the new world of SAS), and 19 (administrator tips).

Checking Out Real-World Success Stories

As users and employees of SAS, we have personally seen many real-world SAS success stories. From forecasting to data warehousing to data mining to business intelligence, SAS can meet just about any need you can imagine. To read a wide array of detailed SAS case studies, use your favorite Web browser and go to `www.sas.com/success`.