SAS Training® Experiences in Our R&D Environment

Maureen Fulcher Mele, Unilever Research U.S., Inc., Edgewater, NJ

ABSTRACT

It started out as a simple question from one scientist, "how can I learn to use SAS® software to analyze data?" That inquiry resulted in the painstaking project of developing and implementing a four day course entitled, "SAS® Software Fundamentals for R&D" for researchers at our site. This paper describes the development of the training course and its impact.

INTRODUCTION

After a scientist requested SAS software training, we issued a site wide survey to determine the interest, by researchers at our R&D facility, in the computer features offered by SAS software. Based upon the survey results, it was decided that a SAS software course tailored to researchers' needs/wants at our R&D facility was needed. It was hoped that the course would be useful in allowing a subject-matter expert to analyze their own data and in alleviating some of the computing tasks performed by our small Statistics Group. To that end, a four day course entitled, "SAS® Software Fundamentals for R&D" was developed. This paper describes the training course and its impact.

Objective:

Train benchtop scientists, engineers, and clinicians in the use of SAS software to allow them to begin to statistically analyze their own experiments.

Considerations:

Course should:
- assume no prior knowledge of SAS software
- be hands-on with numerous workshops and examples
- tailored to our scientists
  - using examples from actual experiments
  - incorporate relevant subject matter.

The generic course offered by SAS Institute had been presented on-site twice previously. Many of the attendees, in the sessions, expressed a desire for a course more tailored to our system and their specific needs.

Prerequisites:

Participants must be familiar with the specific computer operating system where SAS software is utilized including: directory-tree structure, top-level or home directory, creating subdirectories, changing subdirectories, renaming files, copying files, and printing files.

Familiarity with utilizing a text editor (such as nu/TPU®) desired but not essential.

Additional Factor:

After the course development, the computer operating system for SAS software was changed from DEC/VMS® to an open systems environment DEC/Unix®. This resulted in some changes in the course and an additional stumbling block of familiarizing researchers with the new operating system. This was overcome by conducting one-on-one sessions with participants, prior to the SAS software course, to communicate relevant commands.

Design of Course:

Course designed to be 4 full working days, after a 2 day course was unsuccessful (too much material for this time frame).

Booklet containing all overheads and workshops (developed using Freelance® software on PC) provided for all participants at beginning of course. Keymapping and SAS Display Manager System text editor commands provided. OPTIONS for hardcopy printing of graphs to various plotters provided.
Development Time of Course:

The development of the course required approximately five weeks of a single statistician's time. Creation of the overheads in an unfamiliar software package, Freelance, required several more days.

General Overview of Course:

Day 1 - Introduction
   Accessing SAS Software
   Creating a SAS Data Set

Day 2 - Permanent SAS Data Sets
   Data Manipulation
   Combining Data Sets
   Subsetting Data Sets

Day 3 - Introduction to PROCedures
   PROCs For Summary Statistics
   PROCs For Data Sets

Day 4 - Graphics
   Macros

Examples for the course were placed in a computer subdirectory and copied by each participant in the first hands-on workshop.

A detailed outline of the course appears in the Appendix of this paper.

TRAINING

The teaching of the course has been conducted mainly by a single statistician, although it is believed that anyone proficient in SAS software could present the course.

The course has been given as 4 days in succession and as 2 days in 1 week followed by 2 days the following week; the latter is preferred since it allows the participants a chance to tinker with the software on their own and it seems to provide less scheduling problems for participants.

The course size has been limited to a maximum of 8 participants. With a single instructor, a limit of 8 people was set to provide for enough personalized attention during the hands-on workshops.

Workshop solutions have been provided in the booklet, however, participants are urged to complete the workshops on their own. The solutions are provided as a future reference.

For day 4, the graphics section, participants have worked in small groups (or independently as desired). The groups were a necessity under the DEC/VMS computer operating system because of the contention for CPU resources; with the much quicker DEC/Ultrix operating system, this is no longer necessary. However, teamwork is still utilized since it helps to bring some of the struggling participants up-to-speed as well as create a comradery amongst the participants.

RESULTS

The course has been given successfully to 43 participants. In only one instance, a person dropped out after a day and a half because they felt that the course was too overwhelming.

Course evaluations have proved to be very successful. In all cases, the participants rated the course overall as "very good" or "excellent." One point that was noted is that the course has been enjoyed more by individuals who have had prior exposure to SAS software.

We have been successful with people using both display manager to compose and submit programs (this has been possible with the DEC/Ultrix system because of improved performance) as well as via an external text editor. In general, participants who are not familiar with an editor use display manager with some assistance.

In addition, the course has been tailored for specific research departments with particular PROCs added/omitted.

Feedback from course evaluations have been used to continually improve the format, content, and training environment, consistent with our Total Quality Management philosophy.

CONCLUSION

A substantial segment of those trained in this course have become regular users of SAS software and have taken over the day-to-day implementation of
statistical analyses of their data. Those who had been executing SAS software programs (written by others) in the past, now have a better understanding of what they are doing and are capable of modifying, revising, and improving the programs.

FUTURE PLANS

A shortcoming of the course has been its limited concentration on using statistical methods in SAS/STAT® appropriate for analyzing particular types of data. It is planned that a follow-up course will be developed which will discuss statistical methodologies and the implementation of them using SAS software.

NOTICES

SAS, SAS/AF, SAS/ASSIST, SAS/FSP, SAS/GRAPH, SAS/INSIGHT, and SAS/STAT are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Freelance is a registered trademark of Lotus Development Corporation.

tu/TPU is a registered trademark of a/Soft Development, Inc.

VMS and Ultrix are registered trademarks of Digital Equipment Corporation.

Other brand and product names are registered trademarks or trademarks of their respective companies.

AUTHOR

Maureen Fulcher Mele
Unilever Research U.S., Inc. - Edgewater Laboratory
45 River Road
Edgewater, NJ 07020
(201)943-7100
APPENDIX

Detailed Outline of Course:

Day 1

I. Introduction
   A. SAS Development
   B. Operating System, SAS Version
   C. Operating System vs. SAS Commands
      Workshop 1

II. Methods of Executing SAS
   A. Display Manager System (on-line help demo)
   B. Interactive Execution
   C. Batch Execution
   D. SAS/ASSIST®
   E. Other - SAS/Ar® and FSP® Applications
      Workshop 2
   F. Files Created - .log, .lst

III. SAS Data Sets - Backbone to SAS System
   A. DATA vs. PROC Steps
   B. Semicolon
   C. Naming A Data Set
   D. Data Set vs. Variables
   E. Numeric vs. Character Variables

IV. Creating a SAS Data Set And Program
   A. OPTIONS Statement
   B. DATA Statement
   C. INFILE Statement
      INPUT Statement
   D. CARDS Statement
      Comments
   E. PROC SORT/PROC PRINT
      Workshop 3

Day 2

V. SAS Data Libraries
   A. Permanent SAS Data Sets
      - LIBNAME Statement
      - Two-Level Data Set Names
      Workshop 4

VI. Data Manipulation
   A. Calculation of Additional Variables
      - SAS Operators
      - SAS Functions
      - Do Loops
   B. SET Statement
      Workshop 5
C. OUTPUT Statement
D. Combining Data Sets
   - Appending
   - Merging
   - Merging By Variable
E. Subsetting Data Sets
   - IF Statement
   - IF/THEN/ELSE
   - DO Blocks
   - WHERE Statement
F. KEEP/DROP/LENGTH Statements

Workshop 6

REVIEW OF DAYS 1 AND 2
   Workshop 7

Day 3

VII. Intro to PROCs - General Statements
   A. DATA= Option
   B. BY Statement
   C. WHERE Statement
   D. VAR Statement
   E. OUTPUT OUT= Statement
   F. TITLEn Statement

VIII. PROCs For Summary Statistics
   A. PROC SORT
   B. PROC PRINT
      Workshop 8
   C. PROC MEANS
      Workshop 9
   D. PROC FREQ
      Workshop 10
   E. PROC UNIVARIATE
      Workshop 11

IX. PROCs For Data Sets
   A. PROC DATASETS
      Workshop 12
   B. PROC FSEDIT
   C. PROC FSVIEW
      Workshop 13

Day 4

X. Graphics
   A. GOPTIONS Statement
   B. PROC GPLOT
      - PLOT Statement
      - FOOTNOTEn Statement
      - SYMBOLn Statement
      - LEGENDn Statement
      Workshop 14
C. PROC GCHART
   - BLOCK Statement
   - PATTERNn Statement
   Workshop 16
   - VBAR Statement
   - HBAR Statement
   Workshop 17
D. PROC GCONTOUR
   Workshop 18
E. PROC G3D
   Demo
F. PROC GSLIDE
   Demo

XI. Special Demo of the power of SAS/ASSIST and SAS/INSIGHT® for exploring data in an X-windows interface.

XII. Macro Language
   %LET

XIII. Documentation of SAS