CONVERSATIONAL INSTRUCTION IN THE STATISTICAL ANALYSIS SYSTEM

Dorsey A. Glenn and Ione H. Cockrell
Computer Services Division
University of South Carolina
Columbia, South Carolina 29208

ABSTRACT


Teaching relatively unsophisticated users the Statistical Analysis System poses special difficulties in a large geographically dispersed computer network. An interactive SAS course has been developed at the Computer Services Division of the University of South Carolina which overcomes many of these difficulties. Students may sign on to the computer assisted course at any one of approximately 300 interactive terminals throughout the network. The course, which introduces the potential SAS user to the USC computer system and to SAS, consists of textual material interspersed with questions. Each student is given instant feedback to his responses, may see as many or as few of the sessions as he desires, and may completely determine the pace of the course.

PROBLEM STATEMENT

Academic consultants always face the monumental task of training and guiding students, professors, and staff in the use of computer software. The problem at the Computer Services Division of the University of South Carolina is compounded by the size and geographic dispersion of the USC computer network. The network has a 7.0 meg IBM 370/168-3 running under OS/MVS with JES2 supporting 33 batch terminals and VSPC supporting approximately 300 interactive devices. In addition to the main campus, the network serves 8 regional campuses, 5 other state supported colleges, 14 state agencies, and 16 technical education centers.

The Computer Services staff offers a series of short courses at the main campus three times per year to provide the user community with introductions to computer software and information concerning modifications made at CSD-USC. A two week short course (10 class hours) in SAS is a part of this series. The short courses have always been well received, but we have realized that they do not meet the needs of all users, particularly the relatively inexperienced users. Often beginners do not anticipate the need for statistical programs and, therefore, do not take advantage of short courses when they are offered. Since we service such a heterogeneous group, the fixed pace of one short course may be too slow for some and too fast for others. Other problems associated with the short course approach to teaching SAS include scheduling conflicts and the distance to the main campus.

The difficulties and inconveniences associated with the SAS short courses have resulted in a large amount of consultant time being invested in teaching the potential SAS user on a one-to-one basis. The problems associated with this approach are numerous. The amount of staff time taken is excessive; it is impossible for the user to absorb all the necessary information; consultants are often tempted to do the work for the user; and in general, the results of such an approach are highly unsatisfactory.

The statistical consulting staff at CSD-USC decided to investigate alternative training mechanisms which could be used in conjunction with the SAS short courses. We desired a technique which would deal effectively with the user who has little or no background in computers or statistics. The training should not take an excessive amount of user or staff time, but should give a thorough presentation of the basics of SAS. After the training program the user should be knowledgeable enough to utilize consulting sessions effectively. Ideally, the mechanism should also increase the user's confidence with computing equipment.

SOLUTION TO PROBLEM

We decided, after much deliberation, to develop an introductory SAS course called "An Introduction to the Statistical Analysis System (SAS67)" through IBM's Interactive Training System (ITS). ITS employs interactive terminals for communication between the user and the computer during both the authoring and teaching phases of the course. The necessary interactive devices are available for teaching at all locations within the CSD-USC network. This system guides the student through a course by analyzing his response to questions. Any course developed through ITS is divided into sessions which may be composed of pretests, posttests, presentation of text, quizzes, and questions. ITS has features which allow the student to ask for definitions of terms, hints, and even transmit messages to the author.

COURSE CONTENT

The SAS course which has been written at CSD-USC is divided into seven instructional sessions and an introduction to the course. The introduction acquaints the student with the interactive terminal and with the basic structure of the course. The seven instructional sessions are described briefly below.
Session 1: CASES, VARIABLES, AND VALUES is designed for the student with little or no background in computing. It introduces and demonstrates the fundamental concept of case, and distinguishes between numeric and character variables. Specific recommendations are made concerning the "best" way to code and keypunch data, and how to handle multiple cards per case.

Session 2: DATA INPUT CARDS discusses the general rules for SAS statements and names, as well as how to input data. Data cards similar to the ones discussed in Session 1 are used to demonstrate how the DATA, INPUI, and CARDS statements should be used in a SAS program. Column input is used to demonstrate how to read character and numeric data, and how to read multiple records per case.

Session 3: SAS PROCEDURES introduces the concept of a SAS PROC and provides the motivation for each PROC's. The procedures PRINT, PLOT, and MEANS are discussed in detail with various options, parameters, and procedure information statements.

Session 4: JOB CONTROL LANGUAGE is a very brief session intended to introduce the user to basic JCL. Since the course deals with card input only, very little instruction is necessary. The JOB card and the importance of correct spacing are emphasized.

Session 5: PROC SORT AND THE BY STATEMENT presents a short discussion of the procedure SORT and its use with other procedures in conjunction with the BY statement. Students are made aware that this session is not essential to using SAS, but that the technique can be time and effort saving.

Session 6: TRANSFORMATION OF VARIABLES introduces students to simple variable transformations. Assignment statements, IF-THEN statements, and creation of category variables are discussed in detail with numerous questions and examples.

Session 7: DATA SET MODIFICATION considers two types of modification to data sets - changing the number of observations and changing the number of variables. IF and DELETE statements are used with INPUI and SET statements to restrict the number of observations added to a data set. KEEP and DROP statements are used to eliminate unwanted variables.

These seven sessions present the basic concepts necessary to execute simple SAS programs with card input. A reference booklet containing sign-on procedures, figures, examples, and forms, and the major points of each session is provided when the student registers for the course. We decided it was easier to include certain material in the guide than repeat it on the screen several times. This also frees the student from taking notes on the key points of the course.

Each of the seven sessions is composed of three major subdivisions. The introduction gives a brief description of the topics which are included in the session. The student is then given the option of viewing the session or proceeding to the next one.

This approach is especially beneficial when the student has some familiarity with SAS. The following demonstrates how this method was used to introduce Session 5. Please note that all student responses are underlined.

* SESSION 5 *
* PROC SORT and the BY Statement *

In this session we will discuss another SAS procedure called SORT and how it may be used together with a BY statement to enable the user to perform other SAS procedures on many subsets of a SAS data set. For example, a researcher may wish to find means and standard deviations of a set of variables first for his female subjects and then for his male subjects.

By the end of this session you should be able to:
1. use PROC SORT
2. use the BY statement

This session is not necessary to run every SAS job, but it does present material which could save the user a great deal of time and effort. Would you like to see this session? Please answer yes or no.

YES

Fine, let's get started!

The major part of each session consists of tutorial material interspersed with questions. The most common instructional approach is to present one or two concepts with illustrations and examples in a few frames of text and then question the student concerning the material. Frequently, we will anticipate certain wrong answers. These wrong answers may then be used to clarify certain points. Below, we show how this technique was used in Session 1 to help a user understand the concept of a case or observation.

The case or observation is the basic unit of analysis in a survey or experimental study. In a study of student political opinion, the information collected about a particular student, Jerry Lovelace, would represent a case in the analysis. The cases in another study could be industrial output information gathered for each state in the union. Such a study would contain fifty cases.
Forty household surveys were administered in each of sixty American cities. The number of persons in each household, monthly household income, number of persons working 30 or more hours per week, and the number of persons 18 years of age and under were collected in each household.

How many cases does this study have?

Forty surveys were administered in each of sixty cities. Now, how many total surveys does that make?

TWO HUNDRED AND FORTY

You're on the right track, but I think your multiplication is a little off.

2400 +

Very good! The total number of cases would be 60 times 40 or 2400.

The author is never able to anticipate all possible wrong answers. When a student answers with an unexpected response, he may be given a message of the author's choosing and be asked to make another attempt. After a maximum of three unexpected responses, the student may be given the correct answer and an appropriate message.

The following excerpt from Session 2 demonstrates such a situation.

Consider the following input statement:

INPUT NAME $ 1-24 TEST1 27-29
#2 TEST2 3-5;

What type variable is NAME?

ALPHA

No, remember that in Session 1 we discussed two types of variables. Try once more.

HINT

There are two types of variables: character and numeric.

CHARACTER +

(Student responds with almost correct spelling)

Right you are!

(System indicates that he is correct)

Note the use of the HINT option demonstrated in this example. If the student feels he needs some aid in order to answer a question, he may ask for assistance by typing HINT. The user is then given a hint or told that no hints are available. Notice also that in this question, we did not feel it necessary for the student to spell his answer correctly. ITS has a provision which allows the author to accept variations in spelling if he desires. Here, we would accept any spelling which began with "char".

We used true-false, matching, and multiple choice questions in addition to the short answer type questions which have been demonstrated so far. The multiple choice question which follows demonstrates the usefulness of this type of question.

What SAS statement would you use to replace the value of a variable called X with 3 times the sum of X and 15?

A) X = (3X) + 15
B) X = 3*(X + 15)
C) 3*(X + 15) = X
D) X = 3(X + 15)

A

No, this is not even an acceptable SAS statement. Note that 'A' is always needed to indicate multiplication. Therefore, the expression '3X' is not acceptable. Try again.

C

No, the variable which is to be created or whose value is to be replaced is always to the left of the equal sign.

D +

Some students will not be able to absorb all the information after participating in a session for the first time. For this reason, a short summary is given at the end of each session and the student is given the option of repeating the session as many times as he desires. Below we show how this technique was used at the end of Session 3.
In this session, we have discussed, in general terms, the concept of a SAS procedure and how to use such procedures. We also discussed in detail three specific PROC's - FREQ, PRINT, and MEANS. If you do not feel comfortable with this session, you may review it by entering RETURN. If you enter PROCEED, then the next session will be presented.

I'M NOT SURE

Please respond with RETURN or PROCEED.

PROCEED

In that case, you will see Session 4 next. Let's get on with it!

ADVANTAGES OF INTERACTIVE APPROACH

The advantages of this approach are numerous for both the student and the consultant. A few of the advantages for the student are as follows:

1. Learning interactively is more interesting and stimulating for most students than reading a manual or listening to an instructor.
2. Using a terminal increases student confidence with computing equipment.
3. The course may be begun at any time during the year.
4. The student's learning schedule is restricted only by the availability of the terminal.
5. It is not necessary to travel to the main campus to participate.
6. The pace of the course is completely determined by the student since he may sign on and off as will and may repeat or skip sessions as needed.
7. The average student will need only about seven hours to complete the course.

The advantages for the consultant are just as important as those for the student.

1. After the course has been developed, the only staff time which will be required is that to register students and keep the course current.
2. Changes and additions to the course can be made quickly and easily from an interactive terminal.
3. ITS contains facilities which allow the author to evaluate the effectiveness of the course by examining the correct response rate for the questions.

PROBLEMS IN DEVELOPMENT

We are enthusiastic about ITS and the introductory SAS course which we have developed, but we did encounter problems and we anticipate others. Learning ITS required a larger expenditure of time than any other single phase of the development. The ITS manual is somewhat vague and there is no knowledgeable person in our city's IBM office with which to consult directly. We were able, though, to take advantage of a helpful ITS course concerning authoring such courses. We are concerned about slow response times in peak periods of the day after the course is fully implemented, and we anticipate even longer waits as the amount of work done within our system grows.

FUTURE PLANS

Now that the course itself has been finished, we move on to the evaluation phase of our project. We plan to keep the course in a test status for a few months. Individuals who need to learn SAS will be placed in the course and monitored carefully during this test phase. The feedback from this period will be used to evaluate and improve the course. The final version will be publicized and made generally available to all persons within the CSD-USC network. In addition to individuals using our course, several professors have expressed a desire to incorporate our SAS course into their statistics courses.

We feel that our course "An Introduction to the Statistical Analysis System (SAS76)" has great potential for filling a wide variety of different needs both within CSD-USC and the academic community. The response to our course thus far has been highly favorable and encouraging. Despite a few problems and minor setbacks, we feel that an interactive approach to teaching short courses will prove to be highly satisfactory for our computer environment and for others of a similar nature.

REFERENCES