THE REPORT PROCEDURE: AN INTRODUCTION AND OVERVIEW
Ray Pass, Ph. D. - Columbia-Presbyterian Medical Center

The purpose of this paper is to present an overview and introduction for users of the SAS® System (version 6.06) to a new procedure, namely PROC REPORT. As this is a brand new procedure which is still evolving, the timeliness, accuracy and completeness of the information contained in this paper (the written version) is an approximation at best. As of the writing of this paper (February, 1991), SAS Institute has distributed Experimental Release 1 of PROC REPORT to all version 6.06 sites (Experimental Release 2 has been sent to all sites that specifically requested it), with the proviso that the product as released is an experimental version. This proviso is EXTREMELY important; working with either release is a frustrating, but enticing, promise of things to come, only a short way down the road. In fact, the warning you are presently reading may be totally obsolete by the time it is actually published. The bottom line is that PROC REPORT is here and it is going to be a major tool in the SAS System user's report writing arsenal.

If you're hoping to read a little bit of this paper and then quickly scan the rest of it to see if the diagrams and figures appeal to you before reading ALL the words, you can stop now. There are no diagrams and figures in this, the written version. There are two reasons for this. One reason is that PROC REPORT is mainly an interactive procedure, and as such, involves many screen images - too many to include in a written version and still provide a feel for what the procedure can do, and how it is done. The other reason is that the procedure is still evolving, and the version that I used to prepare this paper is not as complete as the current version that the Institute is using, and is certainly not as complete as that which will be eventually released in a production mode. I am aware of changes and additions already, and there will certainly be more before actual release.

WHAT IS PROC REPORT?

A first point to expound on is: What is PROC REPORT and where does it fit in with respect to the other report generating tools provided in the SAS System? PROC REPORT is a report writing procedure which can be used with a set of interactive windows to create and modify (customize) basic PROC PRINT-like output through a process of manipulation of the actual formatted output, not code. You get an immediate view of what the output looks like because you're working directly on it. The concept of interactively "screen painting" the desired output from actual processed results is new to the SAS System.

If a totally vanilla

PROC REPORT DATA=MYDATA;
RUN;

command is issued, the result will be a PROC PRINT look-alike. The similarities end, and the differences begin at that point. Whereas PROC PRINT creates an output that is fixed and cannot be modified without rerunning the procedure from the beginning, and then only allows minor modifications, PROC REPORT basically begins where PROC PRINT ends. The output from a PROC REPORT submission can be sent to the REPORT window instead of the Display Manager OUTPUT window. It can then be modified in numerous ways including customization of headings, rearrangement of columns, addition of new variables computed from existing column variables, changing output FORMATS, establishing "grouping" variables, adding summary statistics, and lastly, creating a PROC TABULATE-like row-by-column tabular display. This customization is continued until the display is in its final desired format. At that point, the unseen report definition created by the system to generate the given display can be saved and reused as needed. Further modification can always be added and the revised format can be saved as a new report, or as a replacement to the original. Data is not stored with the report definition. Therefore, the same report format can be used with different SAS system datasets as long as they all contain the appropriate subset of variables.

WHAT CAN PROC REPORT DO?

PROC REPORT provides many formatting capabilities not available in PROC PRINT (Many of these features are/were/may yet be available in PROC QPRINT, but that's another story which is not appropriate to delve into in depth in this forum. For now, suffice it to say that PROC QPRINT is still very much alive on a daily production basis in too many sites to be overlooked).

Column headings can be defaulted to variable names or labels. Once displayed in the REPORT window, they can be customized as needed. They can be edited in context, or eliminated altogether. They can be split into many lines, can be left, right or center justified, and can be widened or shrunken in width. Headings can be created which span multiple columns.
and can be made to repeat on every page, or to occur only at the beginning or end of the report (footers), or before or after group breaks.

Literal text descriptions, labels, messages, etc. can be placed at various points in the report, and page breaks can be constructed where necessary. Spacing between columns can be adjusted as needed, and can be varied across the page. Output can also be displayed in "telephone book" style where groups of columns are repeated across the page in multiple sets. Formats of variables can be selected as desired, and observations can be made to "wrap" around so that values of all variables for an observation are presented before the next observation is displayed.

Variables can be designated in PROC REPORT as one of six different types. The properties of these different types essentially describe the major construction features of the procedure.

a) DISPLAY - These are variables which make up the standard columns. Values are presented as they appear in the dataset as far as content and order of presentation, one line per observation (or as many lines as are necessary to complete the observation if WRAP is in effect.) Formats for each variable are individually chosen.

b) ORDER - These are "sort" variables and determine the overall order in which observations are presented. There can be multiple ORDER variables, in which case the sort order is left to right (just like PROC SORT.) ORDER variables can be ASCENDING or DESCENDING. ORDER variable values are displayed only when they change, or when the value of an ORDER variable on the left changes. In other words, the display visually branches from left to right.

c) GROUP - These are variables which are used to summarize, group, or compress observations. Summary STATISTICS such as SUM, MEAN, MINIMUM, MAXIMUM, etc. are used with GROUP variables. For example, if SEX and ACADYEAR were used as GROUP variables, a report would have eight lines in its output: (Female-Freshman, Female-Sophomore, Female-Junior, Female Senior, Male-Freshman, Male-Sophomore, Male-Junior, Male-Senior) regardless of how many actual students make up the dataset. ANALYSIS variables (see below) such as GPA and AGE might be used with a STATISTIC of MEAN. The report would then consist of eight lines and four columns (SEX ACADYEAR GPA AGE) and the last two columns would be the mean GPA's and ages for the eight groups ("rows"). GROUP variables also determine the order of presentation of the data, like ORDER variables. GROUP variables must have DISPLAY or ANALYSIS variables and connected STATISTICS associated with them. If they don't, they do not compress or group the observations, and are treated as ORDER variables.

d) ACROSS - These are like GROUP variables except that their values are arranged horizontally across the top. They form a PROC FREQ-like or PROC TABULATE-like array. Using the SEX-ACADYEAR example above, if one of the two variables were used as an ACROSS variable, the report would have either two rows (Female, Male) and five columns (SEX, Freshman, Sophomore, Junior, Senior), or four rows (Freshman, Sophomore, Junior, Senior) and three columns (ACADYEAR, Female, Male), depending on the desired orientation and which of the two variables were used as the ACROSS variable. The values displayed in the "cells" would be determined by the designated ANALYSIS variables present, and the associated STATISTICS. If no ANALYSIS variables were present, frequency counts would be displayed in the cells. ACROSS variables, like GROUP variables, can be nested to form vertical and horizontal intersecting "tree" structures.

e) ANALYSIS - These are summary variables used with GROUP, ACROSS, or ORDER variables. Each ANALYSIS variable must have an associated STATISTIC (the default is SUM). When observations are collapsed, or grouped, according to a GROUP or ACROSS variable, the value shown is that of the summary STATISTIC of the ANALYSIS variable for each set of observations grouped/collapsed together. If ORDER variables are used instead of GROUP variables, each observation is reported individually (no collapsing). If however, an ORDER variable is also defined as a BREAK variable, then summary STATISTICS can be obtained for ANALYSIS variables at each "break" in the observations, determined by the ORDER variable.

f) COMPUTED - These are "variables" that are computed in the REPORT process for display purposes only. They are created during the REPORT design process, in a separate COMPUTE window, by utilizing DATA step statements operating on variables contained in the dataset being reported on, or on other COMPUTE'd variables. They are not contained in the dataset being used, nor are they added to the dataset; they are transient to the REPORT procedure.

The concept of BREAKs mentioned above, is distantly related to the SUMBY and PAGEBY statements found in PROC PRINT, except that its implementation in PROC REPORT is considerably enhanced. Any GROUP or ORDER variable can additionally be designated as a BREAK variable. Summary statistics can then be calculated and displayed, either before or after each break, the breaks being determined by a
change in value of the BREAK variable. The output can be made to start on a new page at each break if desired, and underlines, overlines, skips, and literal text lines can be inserted to visually separate and identify the subsets of the "broken" data. These separating BREAK lines can also be added above or below the whole report.

HOW DOES PROC REPORT WORK?

PROC REPORT can be used to design reports in an interactive manner only, at this time. The SAS System generates the report definition necessary to create the format you "screen paint", and can then be instructed to save this definition in a SAS system catalog. The definition can then be recalled and executed in either an interactive PROC REPORT session, or can be submitted in batch mode by referring directly to the saved catalog entry.

The report definition generated by the system to run a given on-the-fly created formatted output exists in internal 'C' data structures; no SAS system code is generated unless and until a LIST option is used. The LIST option can be issued with the PROC REPORT command in which case the code will appear in the LOG, or it can be invoked via a Display Manager PMENU item wherein it will be displayed in a Display Manager MESSAGE window.

As of this writing, no PROC REPORT batch procedure language has been released, but I'm assured by reliable sources that this will be included in the production release (the new production manual does contain batch language documentation). The result will be that sites that support (or allow) batch processing only (no windowing environments, i.e. Display Manager, PROC REPORT) will still be able to use PROC REPORT via code submitted in batch mode. In addition, the ability to create a report definition interactively, and then have the system convert it to SAS system code, available for further modification, will also be present. Sites that support interactive processing and are not limited to batch mode, offer users a choice: interactive or batch. Windowing capabilities in version 6.06 are much more powerful than those in earlier versions, and even if not used routinely, should be investigated for select purposes, like PROC REPORT.

The PROC REPORT windows are actually independent of Display Manager windows, but they work well with each other. If needed, PROC REPORT can be run interactively in full screen window mode while running the SAS system with the NODMS option specified (Interactive Line Mode). This combination is not preferred, but serves to present the status of PROC REPORT as an independent procedure capable of being invoked in its own interactive windowing environment.

When using PROC REPORT in the interactive design phase, the procedure is implemented with a WD (WINDOWS) option. This opens the REPORT window where the interactive work is done. PROC REPORT also offers a PROMPTER facility when operating in WD mode. This is a step-by-step walkthrough where all aspects of variable selection and description are handled in a question and answer process. First, variables are chosen to be included in the report from a DATASET VARS window. Variable types, formats, widths and labels are then specified in the PROMPTER window for the selected variables. This selection and description process can also be done in a DEFINE window superimposed on the REPORT window, and any work done in the PROMPTER window can be modified in the DEFINE window as well. Additional variables can also be selected and described in the DEFINE window. A series of other windows (BREAK, COMPUTE, ADDING, LOCATION, STATS, etc.) can be called via RKEYS or PMENU items. These are the windows through which most of the design manipulation instructions are passed to the procedure. Column heading manipulations, widths, movements, additions, deletions, etc. are done directly in the REPORT window. As each step is completed in one of these windows, the results are immediately displayed in the REPORT window. RKEYS can be changed via the Display Manager KEYS window and the RKEYS command. When the final acceptable result is achieved via a series of modifications of the displayed output, the underlying report definition can be saved to a catalog, and later recalled for execution or further manipulation. In order to print PROC REPORT output, the procedure must be executed in NOWD mode. The output would then be routed to a printer via the Display Manager OUTPUT or OUTPUT MANAGER window, or via operating system commands if the SAS system has been invoked in NODMS mode. The ability to print PROC REPORT output directly from the REPORT window is planned to be included in the production release.

SHOULD YOU USE PROC REPORT?

Interesting question! You should definitely take a long hard look at it NOW. The product is full of promise, and will undoubtedly be a mainstay when it is released in a full production mode. At this point, working with it is an interesting, enticing, teasing and somewhat frustrating exercise. It is a new way of producing sophisticated and elegant reports with more control than is allowed in PROC PRINT. It seems to match up to PROC QPRINT in prospect and promise,
but the jury of loyal QPRINT users is still out. Rumors have been heard that a conversion program to take existing QPRINT code and translate it to REPORT code is under consideration, if not development. This is needed. Rumors have also been heard of the reinstatement of QPRINT to its rightful position as a fully supported SAS Institute product. This is highly desirable. In the meantime however, take a good long look at PROC REPORT. You will like what you see.

The author may be contacted at:

Columbia-Presbyterian Medical Center
Information Systems Department
161 Ft. Washington Ave. - AP-1310
New York, NY 10032-3784
(212) 305-7748

SAS is a registered trademark of The SAS Institute, Inc., Cary, NC.