

How to Make Data Found in a Print File Available to a Later Step of the Same Program

Hugh Geary, Trilogy Consulting Corporation

Introduction

PROC FREQ, PROC LIFETEST and others produce both the SASdataset* and log file outputs, but do not put as much information in the SASdataset as is put in the printed output. For instance, PROC FREQ did not put Chi-Square and p-values in a SASdataset in versions of SAS prior to 6.07. Also LIFETEST puts Chi-Square values, confidence intervals, etc. in its log file but not in a SASdataset.

However these values are often needed for display in tables and graphs. You can copy these values from the log file by hand into a data set and run a second program which produces the table or graph -OR- you can design the first program to extract the needed values from the log file and pass them on directly to the output part of your program without risking miscopying numbers through human error.

In this paper we examine a means of extracting data from a log file and using it in a later step of the same program that produced the log file. We also look at a program that creates a print file by issuing a system command, SEARCH, and then reads this file in order to process it further. Finally we note that PROC FREQ now produces, in a limited way, an additional SASdataset containing statistics that formerly went only to the log file.

The General Method

- 1) Issue a PRINTTO command to send log output to a file other than the default log file.
- 2) Run the step or procedure which creates the log file.
- 3) Issue another PRINTTO command to send the log output back to the default.
- 4) Read the desired printed file using an infile and input statements and the index command or whatever to identify that is needed. For example, if the character string, 'Chi-Square', is found by the index function and it is known that the Chi-Square value is printed, say, directly beneath or immediately to the right, then the appropriate portion of the line can be incorporated into a variable by using the substring function, and put into numeric form with the input function.
- 5) Finally, (optionally) the 'diverted' log file can be deleted.

A Simple Example

```
filename abc 'aa.lis';

proc printto print=abc new;
data bbb;
  x='This is a test for using PRINTTO.';
  output;
  x="Line two of the same test. &sysdate. &sysptime.";
  output;

proc print data=bbb;
proc printto new;

data _null_;
infile abc length=ln end=lastrec;
input @1 rec $varying132. ln @;
  put rec;

*x "delete/log aa.lis.1";
```

A More Complex Example

Suppose you desire the results of Fisher's Exact Test which is an option of PROC FREQ. The following statement will produce it:

```
proc freq data=xxx;  
  tables y*x /noprint chisq exact;
```

A portion of the PROC FREQ log file is reproduced here:

Statistic	DF	Value	Prob
Chi-Square	4	1.763	0.779
Likelihood Ratio Chi-Square	4	1.785	0.775
MH Chi-Square (Rank Scores)	1	0.365	0.546
Fisher's Exact Test (2-Tail)			0.795
Phi Coefficient		0.111	
Contingency Coefficient		0.110	
Cramer's V		0.111	

Effective Sample Size = 143
Frequency Missing = 5

You must now read this text file locating and retrieving the p-value (or the probability value) found in the right hand column. Each line is read as one character string into a character variable. As each line is read the index function can be used to search for a substring, say, "Fisher" to distinguish this line from all of the other lines in this segment of the log file. To be safe trim trailing blanks off the right end, then convert the last several characters into a number by using the input function.

The following code will do this:
(Suppose the name of the log file is AAA.lis)

```
filename chi_in "AAA.lis";  
  
data abc (keep=pval);  
  infile chi_in length=ln end=last missover;  
  
  input @1 rec $varying132. ln @;  
  
  bbb=index(trim(left(rec)),'Fisher');  
  if bbb>1 then do;  
    pval=input(substr(rec,length(rec)-6),7.3);  
    output;  
  end;
```

You need to know how many Fisher values are in the log file (one per table statement) and the format of the log file. You also need to recheck such code as this when new versions of SAS are brought out.

A more complicated situation arises when several Chi-Square and p-values are needed for one table. Then a series of table statements are needed, producing a number of 'pages' of PROC FREQ output to be read. This is done more easily with a macro.

Reading PROC FREQ Log Files with a Macro

Suppose there are several variables, X1, X2, ... Xn to be compared with the variable, Y. The following code calls two macros to compute the Chi-Square values:

```
%Chi2_A (Dsn_in,X,n,Y,typ=scores-rank);
```

```
%Chi2_B (Dsn_out,Rank);
```

Dsn_in is the name of the SASdataset containing the variables X1,... Xn and Y. The first macro, Chi2_A, creates table statements for each pair, of variables, Xi & Y, i=1,...n. The table statements produce a log file which is processed by the second macro, Chi2_B. Statistics are output in the SASdataset, Dsn_out. If the value of the macro variable, typ, is "scores-rank" then the second argument of the second macro call should be "Rank" - just because the macro was written that way. However, if the value of typ is "exact" then use "2-Tail" for the second argument of the second macro. (This could be further generalized to other Chi-Square options.)

These macros permit processing of log files on a much larger scale. Code for these macros follows:

```
%macro Chi2_A (inn,X,n,Y,typ=);
```

```
***      Compute the Fisher's Exact Test (2-tail) p-values      ***;  
***                               or other p-values                               ***;
```

```
filename abc 'CCC';
```

```
proc printto print=abc new;
```

```
proc freq data=&inn;  
  %do i=1 %to &n;  
    tables &Y*&Xi /noprint chisq &typ.;  
  %end;
```

```
proc printto new;  
%mend Chi2_A;
```

```
%macro Chi2_B (Outt,test);
```

```
*** LOCATE AND RETRIEVE the p-values from THE PROC FREQ OUTPUT PAGE ***;
```

```
filename abc 'CCC';
```

```
data &outt (keep=catg stat type);  
infile abc length=ln end=lastrec missover;  
length stat $7;  
retain catg type prob;  
ff=byte(12);          * printer control character for new page in VMS;
```

```
input @1 rec $varying132. ln @;
```

```
if index(rec,ff)>0 & _n_>10 then link outpt;      *n>7 should suffice;
```

```
*          Pick up the variable name at the end of this title;  
*          Then make catg='X01' instead of 'X1' for sorting later;  
ddd=index(rec,'STATISTICS FOR TABLE');  
if ddd>0 then do;  
  catg=trim(left(substr(REC,length(REC)-2)));  
  catg=substr(catg,1,1) || put(input(substr(catg,2),2.),z2.);  
end;
```

```
*          Picking up the Chi-Square P-value now;  
*          because other selections may fail to exist;  
ddd=index(trim(left(rec)),'Chi-Square');  
if ddd=1 then do;
```

```

type='chisq';
prob=input(substr(REC,length(REC)-6),7.3); *** just CHI-Square ***;
end;

*           Looking for one of the various Chi-Square P-values;
*           such as Fisher's Exact or the Rank-sum, etc.;
ddd=index(rec,"&test");
if ddd>0 then do;
    type="&typ";
    prob=input(substr(REC,length(REC)-6),7.3);
end;

if lastrec then link outpt;
return;

outpt:      if prob=.      then stat=' ';
            else if prob<.001 then stat='<0.001';
*           else if prob<.05  then stat=trim(left(put(prob,6.3)));
            else if prob<=.999 then stat=put(prob,6.3);
                                else stat='>0.999';

            output;
            catg=' '; type=' '; prob=.;
            return;

proc sort data=&outt; by catg;
%mend Chi2_B;

```

Reading a file created by the SEARCH command

Under VMS the command, Search, searches through a specified list of files looking for every occurrence of a specified character string. Every 'found' string is highlighted and the name of the file and the record are listed. The entire collection may be written to a file. For example, the system command,

```
x " search *.sas,*.lis Chisq/out=ccc";
```

issued from within a program, possibly under macro control, creates a file containing all of the references of 'Chisq' in files of the sas type or lis type. ccc may be read as an external file immediately after creation.

New PROC FREQ Output SASdataset Available in SAS version 6.07

Formerly Chi-Square and other statistics were available only in the PROC FREQ log file. But now all of these statistics are available by means of a new Output statement in PROC FREQ. This means that when the Chi-Square or p-values are needed in a later step of the same job they can be gotten directly from a SASdataset instead of having to use methods mentioned above.

This has not happened for any other proc yet; thus, the technique used above will still be needed by procs such as LIFETEST, PHREG and others.

Syntax:

```
PROC FREQ Data=SASdataset
          Order=Internal | freq | data | formatted
          Page
          Noprint; Note: Noprint is new with 6.08 On-line documentation.
```

```
OUTPUT <OUT=SASdataset> <the-output-statistics=list>;
      Note: use the same options here as found in the Tables options.
```

TABLES requests / options;
Weight variable;
By variables;

The options in the Tables statement may be any of the following:

List Missing Out=SASdataset Outpct Chi-Square CMH CMH1 CMH2
Exact Measures PLCORR Totpct and others found in the Procedures Manual.

When the new Output statement is used at least one of the options used in it must be repeated as an option on the Tables statement.

The SASdataset created by it has only one record and only one can be created per invocation of PROC FREQ! Thus PROC FREQ must be called once for each request for which you desire to have statistics output to a SASdataset.

Conclusion

Using PRINTTOs and string functions to grope through external files is still a necessary and expedient way to transfer data from a log file to an easily read SASdataset. Hopefully more of the more frequently used results from procedures will be found in SASdatasets in the future.

*SAS is a trademark or registered trademark of SAS Institute Inc. in the USA and other countries.

Hugh Geary
Trilogy Consulting Corproation
630 Morning Street
Worthington, OH 43085

(614)885-3164