

## Paper 88-27

## Finally – An Easy Way To Compare Two SAS® Files!

Doug Zirbel, Pinnacle Solutions Inc., Indianapolis, IN

**ABSTRACT**

This SAS macro compares two SAS datasets. It produces a differences report which is easy to understand. It does this very simply and it works with either SAS6 or SAS8.

**INTRODUCTION**

Comparing two files to find differences is a common task in the programming world. If the files are text files, or perhaps Microsoft Word files, utility programs exist to do that task.

Here is the main question this macro will answer:

"I have 2 identically-named SAS files, each stored in a different SAS library. They have the same variables and the same variable data types, but the data in each file may be different.

*How do I get a report of the data differences between the two files?"*

The obvious choice for this would seem to be Proc Compare. However Proc Compare's syntax is complex and its output is overkill.

The macro below is a) easy to call and b) produces understandable output.

It finds duplicate records, missing (unmatched) records, and – record-by-record – any differences in variables' values. No, it does not have all the features of Proc Compare, but for simple comparison needs, this will probably do the trick.

Suppose, for example, you wanted to compare the output of a SAS program you wrote with the output of the SAS program after you had made a modification to the code. Save the output file of the first program (e.g., a file called ALPHA) in a SAS library called e.g., FIRSTLIB, and save the output of the second version of the program (also a file called ALPHA) in a library called e.g., SECNDLIB. Then run the macro with the libnames and filename as parameters.

**REQUIREMENTS**

1. Both files must have the **same name**;
2. each file must be in **different SAS libraries** (one could be the default WORK library for example);
3. both files must have the **same set of variables**
4. both files' sets of variables must have identical data types (character, numeric, etc).

**EXPLANATION**

Here is an overview of what happens. An SCL function gets the number of variables in the first input file. Then a macro loop creates one macro variable for each of the variables found in that input dataset (e.g., &X1 for COMPANY, &X2 for ADDR1, &X3 for ADDR2). The file is sorted BY every variable that it has, from the leftmost to the rightmost. Note that the SAS\_\_ALL\_ instruction is used as a shorthand way to do this. The DATA step then identifies any duplicate records it may have, and if they exist they are printed.

The second input file goes through the same steps. After any duplicates have been found, the two sorted files to be compared are then merged by every variable they have. Any record pairs that do not exactly match are put into exception files.

If you want to see output reports with the variables sorted *in a different order* than leftmost-variable-to-rightmost-variable, the two files can be sorted by variables *you* specify in the optional SORTVARS= parameter.

Finally, the exception reports are printed.

**CONCLUSION**

This macro simply compares the data of 2 SAS files which have the same names and same variable names. It is well-suited to many before-and-after testing situations. For more complex comparisons, you should use Proc Compare. The macro code, sample input and sample output are found in the Appendices following the Contact Information.

**ACKNOWLEDGMENTS**

Thanks to SAS Tech Support for the initial SCL code.

**CONTACT INFORMATION**

Your comments and questions are valued and encouraged. Contact the author at:

Doug Zirbel  
Pinnacle Solutions, Inc.  
One North Meridian St., Suite 1010  
Indianapolis, IN 46204  
Work Phone: (317) 423-9143  
Fax: (same)  
Email: [dougl@psiconsultants.com](mailto:dougl@psiconsultants.com)  
Web: <http://www.psiconsultants.com>

## Paper 88-27

## APPENDIX 1: The Macro

```

%MACRO EZCOMPAR(LIB1=, LIB2=, FILE=, SORTVARS=);
*****;
*           E Z C O M P A R                               ;
*   MACRO PARMS:                                         ;
*     LIB1 = LIBNAME OF A SASLIB                          ;
*     LIB2 = LIBNAME OF ANOTHER SASLIB                    ;
*     FILE = FILENAME OF SAS FILE THATS IN BOTH SASLIBS  ;
*     SORTVARS = BY-VARS FOR PRINTING OUTPUT REPORTS     ;
*                                                         ;
*****;
%PUT '== STARTING EZCOMPAR MACRO EXECUTION!';

*****;
*                                                         ;
*   D O   L I B 1   F I L E ;                             ;
*                                                         ;
*****;
%LET DSID=%SYSFUNC(OPEN(&LIB1.&FILE)); /*USE THE SCL OPEN FUNCTION*/
%LET CNT=%SYSFUNC(ATTRN(&DSID,NVARS)); /*SCL ATTRN FUNC FOR # OF VARS*/

*****;
* CREATE A DIFF MACRO VAR FOR EACH VAR IN DATASET.  THE SCL VARNAME
* FUNCTION GETS THIS FROM THE FILE IDENTIFIED BY THE SCL DATASET ID;
*****;
%DO I = 1 %TO &CNT;
  %LET X&I=%SYSFUNC(VARNAME(&DSID,&I));
%END;

%LET RC=%SYSFUNC(CLOSE(&DSID));          /* CLOSE DS */

*****;
* SORT ON ALL VARS ( _ALL_ DOES THIS WITHOUT HAVING TO NAME EACH ONE)
*****;
PROC SORT DATA=&LIB1.&FILE OUT=WORK.LIB1FILE(LABEL="&LIB1.&FILE");
  BY _ALL_;
RUN;

*****;
* FIND ANY DUPS BY LOOKING AT THE FINAL (RIGHTMOST) VARIABLE`S BY-GROUP.
* IF IT`S BOTH THE FIRST AND LAST VALUE OF THAT BY-GROUP, THEN IT`S
* UNIQUE, OTHERWISE IT IS A DUPLICATE;
*****;
DATA WORK.OKRECS
  WORK.DUPS;
  SET WORK.LIB1FILE;
  BY _ALL_;
  IF FIRST.&&X&CNT AND LAST.&&X&CNT THEN OUTPUT WORK.OKRECS;
  ELSE OUTPUT WORK.DUPS;
RUN;

*****;
* PRINT DUPS IF THEY EXIST;
*****;
TITLE "*** &LIB1.&FILE DUPS ***";
PROC PRINT DATA=WORK.DUPS;
RUN;

*****;
*
*   N O W---D O---L I B 2---F I L E ;
*
*****;
%LET DSID=%SYSFUNC(OPEN(&LIB2.&FILE)); /* OPEN DS */
%LET CNT=%SYSFUNC(ATTRN(&DSID,NVARS)); /* # OF VARS IN THE DATASET */

*****;
* CREATE A DIFF MACRO VAR FOR EACH VAR IN DATASET ;
*****;
%DO I = 1 %TO &CNT;
  %LET X&I=%SYSFUNC(VARNAME(&DSID,&I));
%END;

%LET RC=%SYSFUNC(CLOSE(&DSID));          /* CLOSE DS USING SCL FUNCTION */

```

## Paper 88-27

```

*****;
* SORT ON ALL VARS ;
*****;
PROC SORT DATA=&LIB2..&FILE OUT=WORK.LIB2FILE(LABEL="&LIB2..&FILE");
  BY _ALL_;
RUN;

*****;
* FIND ANY DUPS ;
*****;
DATA WORK.OKRECS
  WORK.DUPS;
  SET WORK.LIB2FILE;
  BY _ALL_;
  IF FIRST.&&X&CNT AND LAST.&&X&CNT THEN OUTPUT WORK.OKRECS;
  ELSE OUTPUT WORK.DUPS;
RUN;

*****;
* PRINT DUPS ;
*****;
TITLE "*** &LIB2..&FILE DUPS ***";
PROC PRINT DATA=WORK.DUPS;
RUN;

*****;
* IDENTIFY DIFFERENCES BETWEEN THE 2 SORTED INPUT FILES;
*****;
DATA WORK.A_NOT_B
  WORK.B_NOT_A;
  MERGE WORK.LIB1FILE(IN=A)
        WORK.LIB2FILE(IN=B);
  BY _ALL_;
  IF (A AND NOT B) THEN OUTPUT A_NOT_B;
  ELSE
  IF (B AND NOT A) THEN OUTPUT B_NOT_A;
RUN;

*****;
* IF A SORTVAR PARM WAS ENTERED IN ORDER TO PRINT IN AN ORDER THAT YOU
* CHOOSE (RATHER THAN THE INTERNAL SAS VARIABLE ORDER) -- THEN SORT
* BEFORE PRINTING. YES, THE OUTPUT FILE HAS THE SAME NAME AS THE INPUT;
*****;
%IF &SORTVARS GT %THEN %DO;
  PROC SORT DATA=WORK.A_NOT_B;
    BY &SORTVARS;
  RUN;
%END;

*****;
* IF THERE ARE ANY OBS IN THIS DATASET, PRINT THEM ;
*****;
PROC PRINT DATA=WORK.A_NOT_B;
TITLE1 "*****";
TITLE2 "*** RECS IN &LIB1..&FILE BUT NOT IN &LIB2..&FILE";
TITLE3 "*****";
RUN;

*****;
* IF A SORTVAR PARM WAS ENTERED, THEN SORT BEFORE PRINTING ;
*****;
%IF &SORTVARS GT %THEN %DO;
  PROC SORT DATA=WORK.B_NOT_A;
    BY &SORTVARS;
  RUN;
%END;

*****;
* IF THERE ARE ANY OBS IN THIS DATASET, PRINT THEM ;
*****;
PROC PRINT DATA=WORK.B_NOT_A;
TITLE1 "*****";
TITLE2 "*** RECS IN &LIB2..&FILE BUT NOT IN &LIB1..&FILE";
TITLE3 "*****";
RUN;

PROC CONTENTS DATA=&LIB1..&FILE POSITION;
PROC CONTENTS DATA=&LIB2..&FILE POSITION;
%PUT '=== NOW ENDING EZCOMPAR MACRO EXECUTION';
%MEND EZCOMPAR;

```

Paper 88-27

APPENDIX 2: Sample Input

\*\*\* FIRSTLIB.ALPHA \*\*\* 18:18 Tuesday, December 11, 2001

OBS	COMPANY	ADDRESS1	ADDRESS2	CITY	STATE	ZIP
1	FAMOUS.COM	1785 WESTSIDE PKWY.	SUITE 512	BELLVUE	OH	54321
2	AMERICAN PAN	400 INDUSTRIAL DRIVE		GARY	IN	44321
3	ALBRONS INC	P.O. BOX 1501		HILLVILLE	CA	34450

\*\*\* SECNDLIB.ALPHA \*\*\* 18:18 Tuesday, December 11, 2001

OBS	COMPANY	ADDRESS1	ADDRESS2	CITY	STATE	ZIP
1	FAMOUS.COM	1785 WESTSIDE PKWY.	SUITE 512	BELLVUE	OH	54321
2	AMERICAN PAN	400 INDUSTRIAL DRIVE		GARY	IN	44321
3	AMERICAN PAN	400 INDUSTRIAL DRIVE		GARY	IN	44321
4	ALBRONS INC	P.O. BOX 1500		HILLVILLE	CA	34450

APPENDIX 3: Sample Macro Call

%MACRO EZCOMPAR (LIB1=FIRSTLIB, LIB2=SECNDLIB, FILE=ALPHA, SORTVARS=);

APPENDIX 4: Sample Output

\*\*\* EZCOMPAR MACRO OUTPUT \*\*\* 18:22 Tuesday, December 11, 2001  
 \*\*\* SECNDLIB.ALPHA DUPS \*\*\*

OBS	COMPANY	ADDRESS1	ADDRESS2	CITY	STATE	ZIP
1	AMERICAN PAN	400 INDUSTRIAL DRIVE		GARY	IN	44321
2	AMERICAN PAN	400 INDUSTRIAL DRIVE		GARY	IN	44321

\*\*\* EZCOMPAR MACRO OUTPUT \*\*\* 18:22 Tuesday, December 11, 2001  
 \*\*\*\*\*  
 \*\*\* RECS IN FIRSTLIB.ALPHA BUT NOT IN SECNDLIB.ALPHA  
 \*\*\*\*\*

OBS	COMPANY	ADDRESS1	ADDRESS2	CITY	STATE	ZIP
1	ALBRONS INC	P.O. BOX 1501		HILLVILLE	CA	34450

\*\*\* EZCOMPAR MACRO OUTPUT \*\*\* 18:22 Tuesday, December 11, 2001  
 \*\*\*\*\*  
 \*\*\* RECS IN SECNDLIB.ALPHA BUT NOT IN FIRSTLIB.ALPHA  
 \*\*\*\*\*

OBS	COMPANY	ADDRESS1	ADDRESS2	CITY	STATE	ZIP
1	ALBRONS INC	P.O. BOX 1500		HILLVILLE	CA	34450

## Paper 88-27

\*\*\* EZCOMPAR MACRO OUTPUT \*\*\*

18:22 Tuesday, December 11, 2001

## CONTENTS PROCEDURE

Data Set Name: FIRSTLIB.ALPHA	Observations:	3
Member Type: DATA	Variables:	6
Engine: V609	Indexes:	0
Created: 18:22 Tuesday, December 11, 2001	Observation Length:	77
Last Modified: 18:22 Tuesday, December 11, 2001	Deleted Observations:	0
Protection:	Compressed:	NO
Data Set Type:	Sorted:	NO
Label:		

## -----Engine/Host Dependent Information-----

Data Set Page Size:	27648
Number of Data Set Pages:	1
File Format:	607
First Data Page:	1
Max Obs per Page:	358
Obs in First Data Page:	3
Physical Name:	RMX6657.FIRSTLIB.SASLIB
Release Created:	6.090470
Release Last Modified:	6.090470
Created by:	EZCOMPR
Last Modified by:	EZCOMPR
Subextents:	1
Total Blocks Used:	1

## -----Alphabetic List of Variables and Attributes-----

#	Variable	Type	Len	Pos
1	COMPANY	Char	15	0
2	ADDRESS1	Char	20	15
3	ADDRESS2	Char	20	35
4	CITY	Char	10	55
5	STATE	Char	2	65
6	ZIP	Char	10	67

\*\*\* EZCOMPAR MACRO OUTPUT \*\*\*

18:22 Tuesday, December 11, 2001

## CONTENTS PROCEDURE

Data Set Name: SECNDLIB.ALPHA	Observations:	4
Member Type: DATA	Variables:	6
Engine: V609	Indexes:	0
Created: 18:22 Tuesday, December 11, 2001	Observation Length:	77
Last Modified: 18:22 Tuesday, December 11, 2001	Deleted Observations:	0
Protection:	Compressed:	NO
Data Set Type:	Sorted:	NO
Label:		

## -----Engine/Host Dependent Information-----

Data Set Page Size:	8192
Number of Data Set Pages:	1
File Format:	607
First Data Page:	1
Max Obs per Page:	106
Obs in First Data Page:	4
Physical Name:	SYS01260.T182243.RA000.EZCOMPR.SAS609.H02
Release Created:	6.090470
Release Last Modified:	6.090470
Created by:	EZCOMPR
Last Modified by:	EZCOMPR
Subextents:	1
Total Blocks Used:	2

## -----Alphabetic List of Variables and Attributes-----

#	Variable	Type	Len	Pos
1	COMPANY	Char	15	0
2	ADDRESS1	Char	20	15
3	ADDRESS2	Char	20	35
4	CITY	Char	10	55
5	STATE	Char	2	65
6	ZIP	Char	10	67