

Using PROC COMPARE and PROC CONTENTS
to Compare SAS Data Set Values and Characteristics

Tim Lehman and Ann Lehman

SAS Institute Inc.

Why compare SAS data sets?

- Verify that two files are identical in order to delete duplicate members.
- Verify names and characteristics of variables prior to using the SET, MERGE, or UPDATE statement.
- Compare data values following an application.

The primary function of PROC COMPARE is to compare data values following an application. PROC COMPARE can be used to

- verify data updates by comparing 'before' and 'after' images of a SAS data set
- perform a budget analysis from financial data
- monitor changes in sales, transactions, or other money figures
- compare characteristics of variables between two SAS data sets.

Comparing SAS data sets

Comparisons of interest

- Names of variables in two data sets
- Number of observations in two data sets
- Characteristics of variables in two data sets
- Position of variables in two data sets
- Values of pairs of variables

If these five comparisons all prove equal between two data sets then the SAS files are identical, line for line.

PROC COMPARE compares the values of variables within a single data set or between two data sets and reports the differences.

General syntax:

```
PROC COMPARE options;  
VAR variables;  
WITH variables;  
ID variables;  
BY variables;
```

Selected Options

DATA=	names the data set to be used as the base data set for comparison.
COMPARE=	names the data set to be used as the comparison data set.
OUT=	requests that differences for numeric variables be written to a SAS data set.
STATS	requests the printing of summary statistics for the numeric variable pairs that compare unequally.
NOSUMMARY	suppresses the printing of the summary notes.
NOPRINT	suppresses all printed output (used with OUT= option).
METHOD=	specifies the method for comparison (RELATIVE, PERCENT, or ABSOLUTE).
CRITERION=	specifies the criterion for judging the equality of numeric variables.
ALLOBS	requests that values and differences for all matching observations be printed.

These two data sets each contain a single character variable having the same name.

```
DATA COMP1;
  INFILE RAWDATA1;
  INPUT LINE $CHAR15.;

DATA COMP2;
  INFILE RAWDATA2;
  INPUT LINE $CHAR15.;

PROC PRINT DATA=COMP1;
PROC PRINT DATA=COMP2;
```

11	2	12JAN81	11	<-->	11	2	12JAN81	21
12	2	14JAN81	12		12	2	14JAN81	12
13	2	16JAN81	13		13	2	16JAN81	13
14	2	22JAN81	14		14	2	22JAN81	14
15	2	30JAN81	15	<-->	15	2	30JAN81	19
16	2	02FEB81	16	<-->	16	2	14FEB81	16
17	2	15FEB81	17		17	2	15FEB81	17
18	3	08JAN81	18		18	3	08JAN81	18
19	3	31JAN81	19		19	3	31JAN81	19
20	3	01FEB81	20		20	3	01FEB81	20
21	3	02FEB81	21		21	3	02FEB81	21
22	3	10FEB81	22		22	3	10FEB81	22
23	3	11FEB81	23		23	3	11FEB81	23
24	3	15FEB81	24		24	3	15FEB81	24

OBS	LINE	OBS	LINE
1	1 05JAN81	1	1 05JAN81
2	1 06JAN81	2	1 06JAN81
3	1 08JAN81	3	1 08JAN81
4	1 12JAN81	4	1 12JAN81
5	1 15JAN81	5	1 15JAN81
6	1 22JAN81	6	1 22JAN81
7	1 30JAN81	7	1 30JAN81
8	1 31JAN81	8	1 31JAN81
9	2 08JAN81	9	2 08JAN81
10	2 11JAN81	10	2 11JAN81
11	2 12JAN81	11	2 12JAN81
12	2 14JAN81	12	2 14JAN81
13	2 16JAN81	13	2 16JAN81
14	2 22JAN81	14	2 22JAN81
15	2 30JAN81	15	2 30JAN81
16	2 02FEB81	16	2 14FEB81
17	2 15FEB81	17	2 15FEB81
18	3 08JAN81	18	3 08JAN81
19	3 31JAN81	19	3 31JAN81
20	3 01FEB81	20	3 01FEB81
21	3 02FEB81	21	3 02FEB81
22	3 10FEB81	22	3 10FEB81
23	3 11FEB81	23	3 11FEB81
24	3 15FEB81	24	3 15FEB81

PROC COMPARE produces two reports, a summary and a detail report. The summary report lists

- the number of observations and the number of variables in each data set
- the number of observations and the number of variables compared
- the number of compared variables that were equal
- the number of compared variables that were unequal
- a listing of the number of unequal observations for each variable that had differences (NDIF).

This report is suppressed by the NOSUMMARY option.

The simplest execution of PROC COMPARE compares the values of all variables with the same name in the DATA= and COMPARE= data sets.

```
PROC COMPARE DATA=COMP1 COMPARE=COMP2;
```

The observations are compared on a one-to-one basis by observation number when no BY or ID statement is used.

OBS	LINE	OBS	LINE
1	1 05JAN81	1	1 05JAN81
2	1 06JAN81	2	1 06JAN81
3	1 08JAN81	3	1 08JAN81
4	1 12JAN81	4 <-->	4 1 12JAN81
5	1 15JAN81	5	1 15JAN81
6	1 22JAN81	6	1 22JAN81
7	1 30JAN81	7	1 30JAN81
8	1 31JAN81	8	1 31JAN81
9	2 08JAN81	9	2 08JAN81
10	2 11JAN81	10	2 11JAN81

COMPARE PROCEDURE
COMPARISON OF COMP1 WITH COMP2

COMPARISON SUMMARY

DATASET	OBS	VARIABLES	COMPARISON RESULTS
COMP1	24	1	EQUAL 0
COMP2	24	1	UNEQUAL 1
MATCHED	24	1	TOTAL 1

ALL COMPARISONS UNEQUAL

VARIABLE TYPE LEN NDIF

LINE	CHAR	LEN	NDIF
		15	4

A detail report is produced showing the differences found at each observation for each variable that had differences.

- When no VAR statement is used, a comparison is performed on all variables having the same name in each data set.

- When no ID statement is used, the observation number identifies the observation that has differences.
- When a character variable is compared, all characters are used in the comparison regardless of the variable length, but only the first 20 characters are printed in the report.

COMPARE PROCEDURE
COMPARISON OF COMP1 WITH COMP2

OBS	VARIABLE LINE	COMPARE LINE
4	1 12JAN81 4	1 12JAN81 2
11	2 12JAN81 11	2 12JAN81 21
15	2 30JAN81 15	2 30JAN81 19
16	2 02FEB81 16	2 14FEB81 16

In the following example, the previous two data sets are created again with three variables instead of only one.

```
DATA COMP1;
  INFILE RAWDATA1;
  INPUT CODE DATE $ NUMBER1;
DATA COMP2;
  INFILE RAWDATA2;
  INPUT CODE DATE $ NUM1;

PROC PRINT DATA=COMP1 NOOBS;
PROC PRINT DATA=COMP2 NOOBS;
```

CODE	DATE	NUMBER1	CODE	DATE	NUM1
1	05JAN81	1	1	05JAN81	1
1	06JAN81	2	1	06JAN81	2
1	08JAN81	3	1	08JAN81	3
1	12JAN81	4 <-->	1	12JAN81	2
1	15JAN81	5	1	15JAN81	5
1	22JAN81	6	1	22JAN81	6
1	30JAN81	7	1	30JAN81	7
1	31JAN81	8	1	31JAN81	8
2	08JAN81	9	2	08JAN81	9
2	11JAN81	10	2	11JAN81	10
2	12JAN81	11 <-->	2	12JAN81	21
2	14JAN81	12	2	14JAN81	12
2	16JAN81	13	2	16JAN81	13
2	22JAN81	14	2	22JAN81	14
2	30JAN81	15 <-->	2	30JAN81	19
2	02FEB81	16 <-->	2	14FEB81	16
2	15FEB81	17	2	15FEB81	17
3	08JAN81	18	3	08JAN81	18
3	31JAN81	19	3	31JAN81	19
3	01FEB81	20	3	01FEB81	20
3	02FEB81	21	3	02FEB81	21
3	10FEB81	22	3	10FEB81	22
3	11FEB81	23	3	11FEB81	23
3	15FEB81	24	3	15FEB81	24

A BY statement is used in this example to illustrate PROC COMPARE output when there is BY-group processing.

A VAR statement and a WITH statement are needed to compare variables with different names.

- The variables listed on the VAR statement must be in the DATA= data set.
- The WITH statement lists the corresponding variables in the COMPARE= data set.
- Any variable on the VAR statement without a corresponding WITH variable is assumed to have the same name in the COMPARE= data set.

```
PROC COMPARE DATA=COMP1 COMPARE=COMP2;
  BY CODE;
  VAR NUMBER1 DATE;
  WITH NUM1;
  FORMAT NUMBER1 NUM1 5.1;
```

When numeric variables are compared, differences and percent differences are included on the detail output. By default, numeric values compare unequally if the relative comparison is greater than the default criterion.

COMPARE PROCEDURE
COMPARISON OF COMP1 WITH COMP2

CODE=1

COMPARISON SUMMARY

DATASET	OBS	VARIABLES	COMPARISON RESULTS
COMP1	24	3	EQUAL 1
COMP2	24	3	UNEQUAL 1
MATCHED	8	2	TOTAL 2

COMPARISONS EQUAL

VARIABLE	TYPE	LEN	COMPARE	LEN
DATE	CHAR	8	DATE	8

COMPARE PROCEDURE
COMPARISON OF COMP1 WITH COMP2

CODE=1

COMPARISONS UNEQUAL

VARIABLE	TYPE	LEN	COMPARE	LEN	NDIF
NUMBER1	NUM	8	NUM1	8	1

Suppose x and y are values being compared.

- The default method is METHOD=RELATIVE. Values are judged unequal if $ABS(x-y)/(ABS(x)+ABS(y))/2 > CRITERION$.
- The METHOD=ABSOLUTE comparison judges values to be unequal if $ABS(x-y) > CRITERION$.
- The METHOD=PERCENT comparison judges values to be unequal if $ABS(x-y) / ABS(x)*100 > CRITERION$.
- The default CRITERION is .001 .

The METHOD= option and the CRITERION= option on the procedure statement can be used to specify the method of comparison and the comparison criterion.

COMPARE PROCEDURE
COMPARISON OF COMP1 WITH COMP2

CODE=1

OBS	VARIABLE NUMBER1	COMPARE NUM1	DIFF.	% DIFF
4	4.0	2.0	-2.00000	-50.00000

Comparisons are made within each value of the BY variable.

COMPARE PROCEDURE
COMPARISON OF COMP1 WITH COMP2

CODE=2

COMPARISON SUMMARY

DATASET	OBS	VARIABLES	COMPARE	RESULTS
COMP1	24	3	EQUAL	0
COMP2	24	3	UNEQUAL	2
MATCHED	9	2	TOTAL	2

ALL COMPARISONS UNEQUAL

VARIABLE	TYPE	LEN	COMPARE	LEN	NDIF
NUMBER1	NUM	8	NUM1	8	2
DATE	CHAR	8	DATE	8	1

COMPARE PROCEDURE
COMPARISON OF COMP1 WITH COMP2

CODE=2

OBS	VARIABLE NUMBER1	COMPARE NUM1	DIFF.	% DIFF
3	11.0	21.0	10.00000	90.90909
7	15.0	19.0	4.00000	26.66667

OBS	VARIABLE DATE	COMPARE DATE
8	02FEB81	14FEB81

PROC COMPARE provides the summary output even if no differences were found.

COMPARE PROCEDURE
COMPARISON OF COMP1 WITH COMP2

CODE=3

COMPARISON SUMMARY

DATASET	OBS	VARIABLES	COMPARE	RESULTS
COMP1	24	3	EQUAL	2
COMP2	24	3	UNEQUAL	0
MATCHED	7	2	TOTAL	2

COMPARISONS EQUAL

VARIABLE	TYPE	LEN	COMPARE	LEN
NUMBER1	NUM	8	NUM1	8
DATE	CHAR	8	DATE	8

The ID statement forces a match of the ID variable values.

- If no match is found for the ID variables, no differences are noted.
- The data sets must be sorted by the ID variable within any BY variables used.
- The ID variables must be unique within the data sets.
- The ID variables are included in the detail output.

PROC SORT DATA=COMP1;
BY CODE DATE;
PROC SORT DATA=COMP2;
BY CODE DATE;

PROC COMPARE DATA=COMP1 COMPARE=COMP2 NOSUMMARY;
BY CODE;
ID DATE;
VAR NUMBER1;
WITH NUM1;
FORMAT NUMBER1 NUM1 5.1;

The NOSUMMARY option confines the output to a detail table of differences.

COMPARE PROCEDURE
COMPARISON OF COMP1S WITH COMP2S

CODE=1

DATE	VARIABLE NUMBER1	COMPARE NUM1	DIFF.	% DIFF
12JAN81	4.0	2.0	-2.00000	-50.00000

COMPARE PROCEDURE
COMPARISON OF COMP1S WITH COMP2S

CODE=2

DATE	VARIABLE NUMBER1	COMPARE NUM1	DIFF.	% DIFF
12JAN81	11.0	21.0	10.00000	90.90909
30JAN81	15.0	19.0	4.00000	26.66667

A Simple Application

You want to produce a report that shows how well departments in your company functioned within given budget figures.

Budget figures for each department with each division are recorded in a data set.

```
TITLE 'WONDER PHARMACEUTICAL COMPANY';
DATA BUDGET;
  INFILE SASIN(BUDGET);
  INPUT DIVISION $ DEPT $ BUDGET;
```

```
PROC SORT DATA=BUDGET;
  BY DIVISION DEPT;
```

```
PROC PRINT DATA=BUDGET;
  TITLE2 'DEPARTMENTAL BUDGETS';
  FORMAT BUDGET COMMA10.;
```

WONDER PHARMACEUTICAL COMPANY
DEPARTMENTAL BUDGETS

OBS	DIVISION	DEPT	BUDGET
1	DOMESTIC	ADMIN	28,840
2	DOMESTIC	LEGAL	24,151
3	DOMESTIC	RESEARCH	14,447
4	DOMESTIC	SALES	22,798
5	FOREIGN	ADMIN	28,122
6	FOREIGN	LEGAL	19,212
7	FOREIGN	RESEARCH	25,462
8	FOREIGN	SALES	30,588

The departmental expenditures are recorded each month.

```
DATA EXPENSES;
  INFILE SASIN(EXPENSES);
  INPUT DIVISION $ DEPT $ MONTH $ EXPENSES;
```

```
PROC SORT DATA=EXPENSES;
  BY DIVISION DEPT;
```

```
PROC PRINT DATA=EXPENSES(OBS=15);
  TITLE2 'DEPARTMENTAL MONTHLY EXPENDITURES';
  FORMAT EXPENSES COMMA10.;
```

WONDER PHARMACEUTICAL COMPANY
DEPARTMENTAL MONTHLY EXPENDITURES

OBS	DIVISION	DEPT	MONTH	EXPENSES
1	DOMESTIC	ADMIN	1	180
2	DOMESTIC	ADMIN	2	888
3	DOMESTIC	ADMIN	3	4,832
4	DOMESTIC	ADMIN	4	1,999
5	DOMESTIC	ADMIN	5	1,677
6	DOMESTIC	ADMIN	6	884
7	DOMESTIC	ADMIN	7	5,126
8	DOMESTIC	ADMIN	8	2,921
9	DOMESTIC	ADMIN	9	2,936
10	DOMESTIC	ADMIN	10	4,785
11	DOMESTIC	ADMIN	11	337
12	DOMESTIC	ADMIN	12	4,271
13	DOMESTIC	LEGAL	1	464
14	DOMESTIC	LEGAL	2	5
15	DOMESTIC	LEGAL	3	5,583

Use PROC MEANS to find the total expenditures for each department.

```
PROC MEANS DATA=EXPENSES NOPRINT;
  BY DIVISION DEPT;
  VAR EXPENSES;
  OUTPUT OUT=SPENT SUM=EXPENSES;
```

```
PROC PRINT DATA=SPENT;
  TITLE2 'TOTAL DEPARTMENTAL EXPENDITURES';
  FORMAT EXPENSES COMMA10.;
```

WONDER PHARMACEUTICAL COMPANY
TOTAL DEPARTMENTAL EXPENDITURES

OBS	DIVISION	DEPT	EXPENSES
1	DOMESTIC	ADMIN	30,836
2	DOMESTIC	LEGAL	28,206
3	DOMESTIC	RESEARCH	17,389
4	DOMESTIC	SALES	25,256
5	FOREIGN	ADMIN	26,951
6	FOREIGN	LEGAL	25,381
7	FOREIGN	RESEARCH	21,751
8	FOREIGN	SALES	33,640

PROC COMPARE can be used to compare the output data set from PROC MEANS with the budget figures from the BUDGET data set.

Use PROC COMPARE to compare expenses with budget figures.

- The NOSUMMARY option confines the output to tables that display the difference between budget and expenditures, and the percent of expenditures over or under budget.
- A BY statement causes a comparison to be completed for each division.
- The ID statement forces a match of department within each division, and the department name is used to identify the output.
- The OUT= option is specified to produce an output data set so that differences can be displayed in chart form.

```
TITLE2;
PROC COMPARE DATA=BUDGET COMPARE=SPENT NOSUMMARY
  OUT=TOCHART OUTPERCENT;
  BY DIVISION;
  ID DEPT;
  VAR BUDGET;
  WITH EXPENSES;
  FORMAT BUDGET EXPENSES COMMA6.;
```

The NOPRINT option is often used with the OUT= option to suppress all printed output produced by PROC COMPARE.

The PROC COMPARE report displays the difference between budget and expenditures, and the percent of expenditures over or under budget.

WONDER PHARMACEUTICAL COMPANY
COMPARE PROCEDURE
COMPARISON OF BUDGET WITH SPENT

DIVISION=DOMESTIC

DEPT	VARIABLE BUDGET	COMPARE EXPENSES	DIFF.	% DIFF
ADMIN	28,840	30,836	1996.00	6.92094
LEGAL	24,151	28,206	4055.00	16.79020
RESEARCH	14,447	17,389	2942.00	20.36409
SALES	22,798	25,256	2458.00	10.78165

WONDER PHARMACEUTICAL COMPANY
COMPARE PROCEDURE
COMPARISON OF BUDGET WITH SPENT

DIVISION=FOREIGN

DEPT	VARIABLE BUDGET	COMPARE EXPENSES	DIFF.	% DIFF
ADMIN	28,122	26,951	-1171.00	-4.16400
LEGAL	19,212	25,381	6169.00	32.11014
RESEARCH	25,462	21,751	-3711.00	-14.57466
SALES	30,588	33,640	3052.00	9.97777

The output data set produced by PROC COMPARE contains all BY variables, all ID variables, and all variables listed on the VAR statement.

Because the OUTPERCENT option is used, the value of the variable BUDGET is the percent difference between BUDGET and its corresponding variable EXPENSES, listed on the WITH statement.

```
PROC PRINT DATA=TOCHART;
  TITLE2 'EXPENDITURES AS A PERCENT OVER OR UNDER BUDGET';
  FORMAT BUDGET 6.2;
```

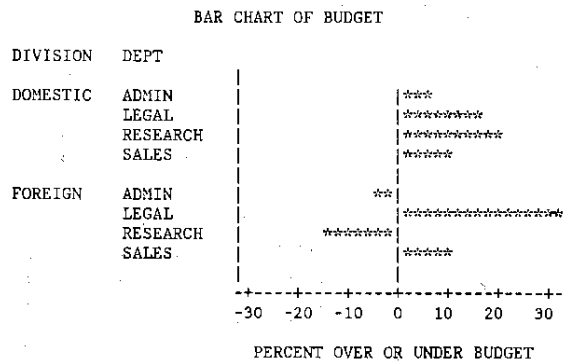
WONDER PHARMACEUTICAL COMPANY
EXPENDITURES AS A PERCENT OVER OR UNDER BUDGET

OBS	DIVISION	DEPT	BUDGET
1	DOMESTIC	ADMIN	6.92
2	DOMESTIC	LEGAL	16.79
3	DOMESTIC	RESEARCH	20.36
4	DOMESTIC	SALES	10.78
5	FOREIGN	ADMIN	-4.16
6	FOREIGN	LEGAL	32.11
7	FOREIGN	RESEARCH	-14.57
8	FOREIGN	SALES	9.98

A good chart is worth a thousand numbers.

```
PROC CHART DATA=TOCHART;
  HBAR DEPT / GROUP=DIVISION SUMVAR=BUDGET
  NOSTATS AXIS= -30 30;
  LABEL BUDGET='PERCENT OVER OR UNDER BUDGET';
  FOOTNOTE 'FISCAL YEAR 1985';
  RUN;
```

WONDER PHARMACEUTICAL COMPANY
EXPENDITURES AS A PERCENT OVER OR UNDER BUDGET



FISCAL YEAR 1985

Other Applications

For applications involving two data sets, it may be necessary to verify that the attributes of the variables are consistent in both data sets.

The differences in the attributes of variables in different data sets need to be known prior to performing operations, such as:

- concatenating
- interleaving
- merging
- updating
- mapping

For variable attribute comparison, you can use PROC COMPARE to compare output data sets created by the OUT= option on PROC CONTENTS.

These two data sets have the same variables and appear to have the same variable attributes.

PROC PRINT DATA=DATA1;

OBS	DEPT	NAME	SEX	SALARY
1	ACCT	John Thompson	M	\$192.00
2	ACCT	Dave Debusscher	M	\$240.00
3	MKT	Patty Allen	F	\$432.00
4	MKT	Jane Tubbs	F	\$233.00

PROC PRINT DATA=DATA2;

OBS	DEPT	NAME	SEX	SALARY
1	ACCT	John Thompson	M	\$234
2	ACCT	Alice Freeman	F	\$241
3	MKT	Alex Trainer	M	\$366

The attributes of these variables should be checked in both data sets prior to operations involving both data sets.

Run PROC CONTENTS on both data sets and use the OUT= option.

```
PROC CONTENTS DATA=DATA1 NOPRINT OUT=CONTENT1;
PROC PRINT DATA=CONTENT1 LABEL;
TITLE 'Listing of PROC CONTENTS Output of Data Set DATA1';
```

Listing of PROC CONTENTS Output of Data Set DATA1

OBS	LIBRARY NAME	MEMBER NAME	DATA SET LABEL	DATA SET TYPE	VARIABLE NAME	VARIABLE TYPE	VARIABLE LENGTH
1	WORK	DATA1			DEPT	2	8
2	WORK	DATA1			NAME	2	20
3	WORK	DATA1			SALARY	1	8
4	WORK	DATA1			SEX	2	8

OBS	LIST 0 NUMBER	VARIABLE LABEL	VARIABLE FORMAT	FORMAT LENGTH	NUMBER OF FORMAT DECIMALS	VARIABLE INFORMAT
1	1	DEPARTMENT		0	0	
2	2			0	0	
3	4		DOLLAR	10	2	
4	3			0	0	

OBS	INFORMAT LENGTH	NUMBER OF INFORMAT DECIMALS	JUSTIFICATION	POSITION IN BUFFER	OBSERVATIONS IN DATA SET
1	0	0	0	4	4
2	0	0	0	12	4
3	0	0	1	40	4
4	0	0	0	32	4

```
PROC CONTENTS DATA=DATA2 NOPRINT OUT=CONTENT2;
PROC PRINT DATA=CONTENT2 LABEL;
TITLE 'Listing of PROC CONTENTS Output of Data Set DATA2';
```

Listing of PROC CONTENTS Output of Data Set DATA2

OBS	LIBRARY NAME	MEMBER NAME	DATA SET LABEL	DATA SET TYPE	VARIABLE NAME	VARIABLE TYPE	VARIABLE LENGTH	LIST 0 NUMBER
1	WORK	DATA2			DEPT	2	8	1
2	WORK	DATA2			NAME	2	15	2
3	WORK	DATA2			SALARY	1	8	4
4	WORK	DATA2			SEX	1	8	3

OBS	VARIABLE LABEL	VARIABLE FORMAT	FORMAT LENGTH	NUMBER OF FORMAT DECIMALS	VARIABLE INFORMAT
1	COMPANY DEPARTMENT		0	0	
2			0	0	
3		DOLLAR	10	0	
4		SEXPMT	1	0	

OBS	INFORMAT LENGTH	NUMBER OF INFORMAT DECIMALS	JUSTIFICATION	POSITION IN BUFFER	OBSERVATIONS IN DATA SET
1	0	0	0	4	3
2	0	0	0	12	3
3	0	0	1	35	3
4	0	0	1	27	3

A visual comparison of the results is tedious. PROC COMPARE can be used to identify and summarize different variable attributes.

The output data set from PROC CONTENTS contains one observation for each variable contained in the input data set. Each variable in the output data set contains attribute information from the input data set.

A listing of the output variables created by PROC CONTENTS is helpful in understanding a PROC CONTENTS output data set.

Name	Type	Description
FORMAT	CHAR	Variable Format
FORMATD	NUM	Number Of Format Decimals
FORMATL	NUM	Format Length
INFORMAT	CHAR	Variable Informat
INFORMD	NUM	Number Of Informat Decimals
INFORML	NUM	Informat Length
JUST	NUM	Justification
LABEL	CHAR	Variable Label
LENGTH	NUM	Variable Length
LIBNAME	CHAR	Library Name
MEMLABEL	CHAR	Data Set Label

```
MEMNAME  CHAR  Member Name
MEMTYPE  CHAR  Data Set Type
NAME     CHAR  Variable Name
NOBS     NUM   Observations In Data Set
NPOS     NUM   Position In Buffer
TYPE     NUM   Variable Type
VARO     NUM   List 0 Number
```

PROC COMPARE Output (continued)

Comparison of Attributes of Data Sets DATA1 and DATA2

COMPARE PROCEDURE
COMPARISON OF CONTENT1 WITH CONTENT2

NAME	VARIABLE TYPE		DIFF.	% DIFF
	VARIABLE TYPE	COMPARE TYPE		
SEX	2	1	-1.00000	-50.00000

Use PROC COMPARE to summarize the differences between the two PROC CONTENTS output data sets.

```
PROC COMPARE DATA=CONTENT1 COMPARE=CONTENT2;
  VAR TYPE LENGTH LABEL
  FORMAT FORMATL FORMATD INFORMAT INFORML INFORMD;
  ID NAME;
  TITLE 'Comparison of Attributes of Data Sets DATA1 and DATA2';
  FORMAT TYPE LENGTH FORMATD FORMATL INFORMD INFORML 1.;
RUN;
```

NAME	VARIABLE LENGTH		DIFF.	% DIFF
	VARIABLE LENGTH	COMPARE LENGTH		
NAME	20	15	-5.00000	-25.00000

In this example, only selected variable attributes are compared.

PROC COMPARE Output

The first listing shows which variable attributes are the same in both data sets.

Comparison of Attributes of Data Sets DATA1 and DATA2

COMPARE PROCEDURE
COMPARISON OF CONTENT1 WITH CONTENT2

COMPARISON SUMMARY

DATASET	OBS	VARIABLES	COMPARISON RESULTS
CONTENT1	4	18	EQUAL 3
CONTENT2	4	18	UNEQUAL 6
MATCHED	4	18	TOTAL 9

COMPARISONS EQUAL

VARIABLE	TYPE	LEN	LABEL
INFORMAT	CHAR	8	VARIABLE INFORMAT
INFORML	NUM	8	INFORMAT LENGTH
INFORMD	NUM	8	NUMBER OF INFORMAT DECIMALS

PROC COMPARE Output (continued)

The attributes that are not the same are listed in the following tables.

Comparison of Attributes of Data Sets DATA1 and DATA2

COMPARE PROCEDURE
COMPARISON OF CONTENT1 WITH CONTENT2

COMPARISONS UNEQUAL

VARIABLE	TYPE	LEN	LABEL	NDIF
TYPE	NUM	8	VARIABLE TYPE	1
LENGTH	NUM	8	VARIABLE LENGTH	1
LABEL	CHAR	40	VARIABLE LABEL	1
FORMAT	CHAR	8	VARIABLE FORMAT	1
FORMATL	NUM	8	FORMAT LENGTH	1
FORMATD	NUM	8	NUMBER OF FORMAT DECIMALS	1

PROC COMPARE Output (continued)

Comparison of Attributes of Data Sets DATA1 and DATA2

COMPARE PROCEDURE
COMPARISON OF CONTENT1 WITH CONTENT2

NAME	VARIABLE FORMAT		COMPARE LABEL
	VARIABLE LABEL	COMPARE LABEL	
SEX	SEX	SEXFMT	

NAME	FORMAT LENGTH		DIFF.	% DIFF
	VARIABLE FORMATL	COMPARE FORMATL		
SEX	0	1	1.00000	

NAME	NUMBER OF FORMAT DECIMALS		DIFF.	% DIFF
	VARIABLE FORMATD	COMPARE FORMATD		
SALARY	2	0	-2.00000	-100.00

Summary

Comparing SAS data sets

PROC COMPARE is a simple SAS procedure that performs comparisons of SAS variable values. Using PROC COMPARE in conjunction with the output data set generated by PROC CONTENTS provides a comparison of data set variables present and data set variable attributes.

COMPARISON TO BE MADE	PROCEDURES
Values of pairs of variables	PROC COMPARE
Names of variables	PROC CONTENTS PROC COMPARE
Number of observations	PROC CONTENTS PROC COMPARE
Characteristics of variables	PROC CONTENTS PROC COMPARE
Position of variables	PROC CONTENTS PROC COMPARE

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