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 COMPl;
INFILE RAWOATA1;
INPUT LINE $CHAR15.;
DATA COMP2;
INFILE RAWDATA2;
INPUT LINE $CHAR15.;
PROC PRINT DATA=COMPl;
PROC PRINT DATA=COMP2;

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=COMPl COMPARE=COMP2;

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

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.

COMPARE PROCEDURE
COMPARISON OF COMPl WITH COMP2

COMPARISON SUMMARY

<table>
<thead>
<tr>
<th>DATASET</th>
<th>OBS</th>
<th>VARIABLES</th>
<th>COMPARISON RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMPl</td>
<td>24</td>
<td>1</td>
<td>EQUAL</td>
</tr>
<tr>
<td>COMP2</td>
<td>24</td>
<td>1</td>
<td>UNEQUAL</td>
</tr>
<tr>
<td>MATCHED</td>
<td>24</td>
<td>1</td>
<td>TOTAL</td>
</tr>
</tbody>
</table>

ALL COMPARISONS UNEQUAL

VARIABLE TYPE LEN NDIF

<table>
<thead>
<tr>
<th>LINE</th>
<th>CHAR</th>
<th>15</th>
<th>4</th>
</tr>
</thead>
</table>

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**

<table>
<thead>
<tr>
<th>OBS</th>
<th>VARIABLE</th>
<th>COMPARE</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1 12JAN81</td>
<td>1 12JAN81</td>
</tr>
<tr>
<td>11</td>
<td>2 12JAN81</td>
<td>1 12JAN81</td>
</tr>
<tr>
<td>15</td>
<td>2 30JAN81</td>
<td>2 30JAN81</td>
</tr>
<tr>
<td>16</td>
<td>2 02FEB81</td>
<td>2 14FEB81</td>
</tr>
</tbody>
</table>

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

**DATA COMP1:**

```plaintext
DATA COMP1;
INFILE RAWDATA1;
INPUT CODE DATE $ NUMBER1;
```

**DATA COMP2:**

```plaintext
DATA COMP2;
INFILE RAWDATA2;
INPUT CODE DATE $ NUM1;
```

**PROC PRINT DATA=COMP1 NOOBS;**

**PROC PRINT DATA=COMP2 NOOBS;**

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**

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

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.

**COMPARE PROCEDURE**

**COMPARISON OF COMP1 WITH 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**

```plaintext
CODE=1
```

**COMPARISON SUMMARY**

```plaintext
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

```plaintext
DATE CHAR 8 DATE 8
```

**COMPARE PROCEDURE**

**COMPARISON OF COMP1 WITH COMP2**

```plaintext
CODE=1
```

**COMPARISONS UNEQUAL**

VARIABLE TYPE LEN COMPARE LEN NDIF

```plaintext
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 \( \frac{\text{ABS}(x-y)}{(\text{ABS}(x)\times\text{ABS}(y))/2} > \text{CRITERION} \).
- The METHOD=ABSOLUTE comparison judges values to be unequal if \( \text{ABS}(x-y) > \text{CRITERION} \).
- The METHOD=PERCENT comparison judges values to be unequal if \( \text{ABS}(x-y)/\text{ABS}(x) \times 100 > \text{CRITERION} \).
- 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.

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>COMPARE</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBS</td>
<td>DATE</td>
</tr>
<tr>
<td>4</td>
<td>02FEB81</td>
</tr>
<tr>
<td>6</td>
<td>14FEB81</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>DATASET</th>
<th>OBS</th>
<th>VARIABLES</th>
<th>COMPARISON RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMP1</td>
<td>24</td>
<td>3</td>
<td>EQUAL</td>
</tr>
<tr>
<td>COMP2</td>
<td>24</td>
<td>3</td>
<td>UNEQUAL</td>
</tr>
<tr>
<td>MATCHED</td>
<td>7</td>
<td>2</td>
<td>TOTAL</td>
</tr>
</tbody>
</table>

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.

\[
\text{COMPARE PROCEDURE} \\
\text{COMPARISON OF COMP1S WITH COMP2S} \\
\text{CODE = 1} \\
\begin{array}{|c|c|c|c|c|} 
\hline 
\text{DATE} & \text{VARIABLE} & \text{COMPARE} & \text{NUM2} & \text{DIFF} & \% \text{DIFF} \\
\hline 
12JAN81 & 4.0 & 2.0 & 2.00000 & 50.00000 \\
\hline 
\end{array}
\]

\[
\text{COMPARE PROCEDURE} \\
\text{COMPARISON OF COMP1S WITH COMP2S} \\
\text{CODE = 2} \\
\begin{array}{|c|c|c|c|c|} 
\hline 
\text{DATE} & \text{VARIABLE} & \text{COMPARE} & \text{NUM2} & \text{DIFF} & \% \text{DIFF} \\
\hline 
12JAN81 & 11.0 & 21.0 & 10.00000 & 90.90909 \\
30JAN81 & 15.0 & 19.0 & 4.00000 & 26.66667 \\
\hline 
\end{array}
\]

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.

\[
\text{TITLE 'WONDER PHARMACEUTICAL COMPANY'} \\
\text{DATA BUDGET;} \\
\text{INFILE SASIN(BUDGET);} \\
\text{INPUT DIVISION $ DEPT $ BUDGET;} \\
\text{PROC SORT DATA=BUDGET;} \\
\text{BY DIVISION DEPT;} \\
\text{PROC PRINT DATA=BUDGET;} \\
\text{TITLE2 'DEPARTMENTAL BUDGETS'} ; \\
\text{FORMAT BUDGET COMMA10.1;} \\
\]

\[
\text{WONDER PHARMACEUTICAL COMPANY} \\
\text{DEPARTMENTAL BUDGETS} \\
\begin{array}{|c|c|c|c|} 
\hline 
\text{OBS} & \text{DIVISION} & \text{DEPT} & \text{BUDGET} \\
\hline 
1 & \text{DOMESTIC ADMIN} & 28,840 \\
2 & \text{DOMESTIC LEGAL} & 24,151 \\
3 & \text{DOMESTIC RESEARCH} & 14,447 \\
4 & \text{DOMESTIC SALES} & 22,795 \\
5 & \text{FOREIGN ADMIN} & 28,122 \\
6 & \text{FOREIGN LEGAL} & 19,212 \\
7 & \text{FOREIGN RESEARCH} & 25,662 \\
8 & \text{FOREIGN SALES} & 30,588 \\
\hline 
\end{array}
\]

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

\[
\text{PROC MEANS DATA=EXPENSES NODRINT;} \\
\text{BY DIVISION DEPT;} \\
\text{VAR EXPENSES;} \\
\text{OUTPUT OUT=SPENT SUUM=EXPENSES;} \\
\text{PROC PRINT DATA=SPENT;} \\
\text{TITLE2 'TOTAL DEPARTMENTAL EXPENDITURES';} \\
\text{FORMAT EXPENSES COMMA10.1;} \\
\]

\[
\text{WONDER PHARMACEUTICAL COMPANY} \\
\text{TOTAL DEPARTMENTAL EXPENDITURES} \\
\begin{array}{|c|c|c|c|} 
\hline 
\text{OBS} & \text{DIVISION} & \text{DEPT} & \text{EXPENSES} \\
\hline 
1 & \text{DOMESTIC ADMIN} & 180 \\
2 & \text{DOMESTIC ADMIN} & 888 \\
3 & \text{DOMESTIC ADMIN} & 4,832 \\
4 & \text{DOMESTIC ADMIN} & 1,999 \\
5 & \text{DOMESTIC ADMIN} & 1,677 \\
6 & \text{DOMESTIC ADMIN} & 884 \\
7 & \text{DOMESTIC ADMIN} & 5,126 \\
8 & \text{DOMESTIC ADMIN} & 2,921 \\
9 & \text{DOMESTIC ADMIN} & 2,936 \\
10 & \text{DOMESTIC ADMIN} & 4,785 \\
11 & \text{DOMESTIC ADMIN} & 337 \\
12 & \text{DOMESTIC ADMIN} & 4,271 \\
13 & \text{DOMESTIC LEGAL} & 464 \\
14 & \text{DOMESTIC LEGAL} & 5 \\
15 & \text{DOMESTIC LEGAL} & 5,583 \\
\hline 
\end{array}
\]

The departmental expenditures are recorded each month.

\[
\text{DATA EXPENSES;} \\
\text{INFILE SASIN(EXPENSES);} \\
\text{INPUT DIVISION $ DEPT $ MONTH $ EXPENSES;} \\
\text{PROC SORT DATA=EXPENSES;} \\
\text{BY DIVISION DEPT;} \\
\text{PROC PRINT DATA=EXPENSES(OBS=15);} \\
\text{TITLE2 'DEPARTMENTAL MONTHLY EXPENDITURES';} \\
\text{FORMAT EXPENSES COMMA10.1;} \\
\]

\[
\text{WONDER PHARMACEUTICAL COMPANY} \\
\text{DEPARTMENTAL MONTHLY EXPENDITURES} \\
\begin{array}{|c|c|c|c|c|} 
\hline 
\text{OBS} & \text{DIVISION} & \text{DEPT} & \text{MONTH} & \text{EXPENSES} \\
\hline 
1 & \text{DOMESTIC ADMIN} & 1 & 180 \\
2 & \text{DOMESTIC ADMIN} & 2 & 888 \\
3 & \text{DOMESTIC ADMIN} & 3 & 4,832 \\
4 & \text{DOMESTIC ADMIN} & 4 & 1,999 \\
5 & \text{DOMESTIC ADMIN} & 5 & 1,677 \\
6 & \text{DOMESTIC ADMIN} & 6 & 884 \\
7 & \text{DOMESTIC ADMIN} & 7 & 5,126 \\
8 & \text{DOMESTIC ADMIN} & 8 & 2,921 \\
9 & \text{DOMESTIC ADMIN} & 9 & 2,936 \\
10 & \text{DOMESTIC ADMIN} & 10 & 4,785 \\
11 & \text{DOMESTIC ADMIN} & 11 & 337 \\
12 & \text{DOMESTIC ADMIN} & 12 & 4,271 \\
13 & \text{DOMESTIC LEGAL} & 1 & 464 \\
14 & \text{DOMESTIC LEGAL} & 2 & 5 \\
15 & \text{DOMESTIC LEGAL} & 3 & 5,583 \\
\hline 
\end{array}
\]

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 COMM6.;
```

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
```

<table>
<thead>
<tr>
<th>DEPT</th>
<th>BUDGET</th>
<th>EXPENSES</th>
<th>DIFF.</th>
<th>% DIFF</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADMIN</td>
<td>28,122</td>
<td>26,951</td>
<td>-1171</td>
<td>-4.16</td>
</tr>
<tr>
<td>LEGAL</td>
<td>19,212</td>
<td>25,361</td>
<td>6149</td>
<td>32.11</td>
</tr>
<tr>
<td>RESEARCH</td>
<td>25,462</td>
<td>21,751</td>
<td>-3711</td>
<td>-14.57</td>
</tr>
<tr>
<td>SALES</td>
<td>30,588</td>
<td>33,640</td>
<td>3052</td>
<td>9.98</td>
</tr>
</tbody>
</table>

```
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;
```

A good chart is worth a thousand numbers.

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

```
WONDER PHARMACEUTICAL COMPANY 
EXPENDITURES AS A PERCENT OVER OR UNDER BUDGET 
DIVISION=FOREIGN
```

<table>
<thead>
<tr>
<th>DEPT</th>
<th>BUDGET</th>
<th>EXPENSES</th>
<th>DIFF.</th>
<th>% DIFF</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADMIN</td>
<td>28,122</td>
<td>26,951</td>
<td>-1171</td>
<td>-4.16</td>
</tr>
<tr>
<td>LEGAL</td>
<td>19,212</td>
<td>25,361</td>
<td>6149</td>
<td>32.11</td>
</tr>
<tr>
<td>RESEARCH</td>
<td>25,462</td>
<td>21,751</td>
<td>-3711</td>
<td>-14.57</td>
</tr>
<tr>
<td>SALES</td>
<td>30,588</td>
<td>33,640</td>
<td>3052</td>
<td>9.98</td>
</tr>
</tbody>
</table>

```
WONDER PHARMACEUTICAL COMPANY 
COMPARE PROCEDURE 
COMPARISON OF BUDGET WITH SPENT 
DIVISION=FOREIGN
```

```
WONDER PHARMACEUTICAL COMPANY 
EXPENDITURES AS A PERCENT OVER OR UNDER BUDGET 
BAR CHART OF BUDGET 
DIVISION=FOREIGN 
```

```
WONDER PHARMACEUTICAL COMPANY 
EXPENDITURES AS A PERCENT OVER OR UNDER BUDGET 
BAR CHART OF BUDGET 
DIVISION=FOREIGN 
```

```
FISCAL YEAR 1985
```

936
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;**

<table>
<thead>
<tr>
<th>OBS</th>
<th>DEPT</th>
<th>NAME</th>
<th>SEX</th>
<th>SALARY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ACCT</td>
<td>John Thompson</td>
<td>N</td>
<td>$192.00</td>
</tr>
<tr>
<td>2</td>
<td>ACCT</td>
<td>Dave Babushis</td>
<td>M</td>
<td>$240.00</td>
</tr>
<tr>
<td>3</td>
<td>MKT</td>
<td>Patty Allen</td>
<td>F</td>
<td>$432.00</td>
</tr>
<tr>
<td>4</td>
<td>MKT</td>
<td>Jane Tubbs</td>
<td>F</td>
<td>$233.00</td>
</tr>
</tbody>
</table>

**PROC PRINT DATA=DATA2;**

<table>
<thead>
<tr>
<th>OBS</th>
<th>DEPT</th>
<th>NAME</th>
<th>SEX</th>
<th>SALARY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ACCT</td>
<td>John Thompson</td>
<td>M</td>
<td>$234.00</td>
</tr>
<tr>
<td>2</td>
<td>ACCT</td>
<td>Alice Freeman</td>
<td>F</td>
<td>$261.00</td>
</tr>
<tr>
<td>3</td>
<td>MKT</td>
<td>Alex Trainer</td>
<td>M</td>
<td>$366.00</td>
</tr>
</tbody>
</table>

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;**

**PROC CONTENTS DATA=DATA2 NOPRINT OUT=CONTENT2;**

**PROC PRINT DATA=CONTENT2 LABEL;**

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 the PROC CONTENTS output data set.
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 FORMAT1 FORMAT2 FORMAT3 FORMAT4 FORMAT5 FORMAT6 FORMAT7 FORMAT8 FORMAT9 FORMAT10;
    ID NAME;
    TITLE 'Comparison of Attributes of Data Sets DATA1 and DATA2';
    CONTENT TYPE CONTENT1 CONTENT2;
    RUN;

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.

| COMPARE SUMMARY |
|-----------------|------------------|
| DATASET         | OBS  | VARIABLES | COMPARISON RESULTS |
| CONTENT1        | 4    | 18         | EQUAL              |
| CONTENT2        | 4    | 16         | UNEQUAL            |
| MATCHED         | 4    | 16         | TOTAL              |
|                  |      |            | 9                  |

PROC COMPARE Output (continued)
The attributes that are not the same are listed in the following tables.

<table>
<thead>
<tr>
<th>COMPARISONS EQUAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>VARIABLE TYPE</td>
</tr>
<tr>
<td>LEN</td>
</tr>
<tr>
<td>LABEL</td>
</tr>
<tr>
<td>INFORMAT CHAR</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>VARIABLE INFORMAT</td>
</tr>
<tr>
<td>INFORMAT NUM</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>INFORMLAT</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>NUMBER OF INFORMAT DECIMALS</td>
</tr>
</tbody>
</table>

PROC COMPARE Output (continued)
Comparison of Attributes of Data Sets DATA1 and DATA2

<table>
<thead>
<tr>
<th>COMPARE PROCEDURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMPARISON OF CONTENT1 WITH CONTENT2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VARIABLE TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAME</td>
</tr>
<tr>
<td>COMPARE TYPE</td>
</tr>
<tr>
<td>DIFF.</td>
</tr>
<tr>
<td>SEX</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VARIABLE LENGTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAME</td>
</tr>
<tr>
<td>COMPARE LENGTH</td>
</tr>
<tr>
<td>DIFF.</td>
</tr>
<tr>
<td>NAME</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VARIABLE LABEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAME</td>
</tr>
<tr>
<td>COMPARE LABEL</td>
</tr>
<tr>
<td>NAME</td>
</tr>
<tr>
<td>COMPANY DEPARTMENT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VARIABLE NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMPARE NAME</td>
</tr>
<tr>
<td>NAME</td>
</tr>
<tr>
<td>SEX</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VARIABLE FORMAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAME</td>
</tr>
<tr>
<td>COMPARE FORMAT</td>
</tr>
<tr>
<td>NAME</td>
</tr>
<tr>
<td>SEX</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NUMBER OF FORMAT DECIMALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAME</td>
</tr>
<tr>
<td>COMPARE NUMBER OF FORMAT DECIMALS</td>
</tr>
<tr>
<td>NAME</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>COMPARISON TO BE MADE</th>
<th>PROCEDURES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Values of pairs of variables</td>
<td>PROC COMPARE</td>
</tr>
<tr>
<td>Names of variables</td>
<td>PROC CONTENTS</td>
</tr>
<tr>
<td>Number of observations</td>
<td>PROC COMPARE</td>
</tr>
<tr>
<td>Characteristics of variables</td>
<td>PROC CONTENTS</td>
</tr>
<tr>
<td>Position of variables</td>
<td>PROC COMPARE</td>
</tr>
</tbody>
</table>

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