

Jane Eslinger Bio

Jane Eslinger is an independent programmer based in North Carolina with over 20 years of experience with SAS.

Jane has authored two books: The SAS® Programmer's PROC REPORT Handbook: Basic to Advanced Reporting Techniques and The SAS® Programmer's PROC REPORT Handbook: ODS Companion. Her SAS certifications include Advanced Programmer for SAS®9, SAS® Certified Data Scientist, and SAS® Certified Advanced Visual Business Analyst.

Making a Readable PROC COMPARE Report in Excel

Jane Eslinger

Eslinger Enterprises, LLC

Default PROC COMPARE Output

The COMPARE Procedure
Comparison of WORK.BASE with WORK.COMPARE
(Method=RELATIVE(2.22E-09), Criterion=0.00001)

Data Set Summary

Dataset	Created	Modified	NVar	NObs
WORK.BASE	14AUG25:13:43:22	14AUG25:13:43:22	7	1587
WORK.COMPARE	14AUG25:13:43:22	14AUG25:13:43:22	7	1587

Variables Summary

Number of Variables in Common: 7.
Number of Variables with Differing Attributes: 2.

A					
1	The COMPARE Procedure				
2	Comparison of WORK.BASE with WORK.COMPARE				
3	(Method=RELATIVE(2.22E-09), Criterion=0.00001)				
4					
5	Data Set Summary				
6					
7	Dataset	Created	Modified	NVar	NObs
8					
9	WORK.BASE	14AUG25:13:43:22	14AUG25:13:43:22	7	1587
10	WORK.COMPARE	14AUG25:13:43:22	14AUG25:13:43:22	7	1587
11					
12					
13	Variables Summary				
14					
15	Number of Variables in Common: 7.				
16	Number of Variables with Differing Attributes: 2.				

Default PROC COMPARE Output

Listing of Common Variables with Differing Attributes

Variable	Dataset	Type	Length	Format	Label
Area	WORK.BASE	Char	14		Geographical Area
	WORK.COMPARE	Char	14		
Fahrenheit	WORK.BASE	Num	8		
	WORK.COMPARE	Num	8	8.1	

A					
1					
2					
3	Listing of Common Variables with Differing Attributes				
4					
5	Variable	Dataset	Type	Length	Format Label
6					
7	Area	WORK.BASE	Char	14	Geographical Area
8		WORK.COMPARE	Char	14	
9	Fahrenheit	WORK.BASE	Num	8	
10		WORK.COMPARE	Num	8 8.1	

Default PROC COMPARE Output

Observation Summary

Observation	Base	Compare
First Obs	1	1
First Unequal	14	14
Last Unequal	1400	1400
Last Obs	1587	1587

Observation Summary

Observation	Base	Compare
First Obs	1	1
First Unequal	14	14
Last Unequal	1400	1400
Last Obs	1587	1587

Default PROC COMPARE Output

Value Comparison Results for Variables

Obs		Base Value	Compare Value
		Name	Name
		_____ +	_____ +
88		Hot Springs Near Icy	Hot Springs Near Icy

Obs		Base	Compare	Diff.	% Diff
		Celsius	Celsius		
14		154.0000	120.0000	-34.0000	-22.0779
1400		130.0000	120.0000	-10.0000	-7.6923

A					
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					

PROC COMPARE ODS Table Names

- **CompareDatasets** - Information about the data set or data sets
- **CompareVariables** - A listing of differences in variable types or attributes between the base data set and the compare data set
- **CompareSummary** - Summary report of observations, values, and variables with unequal values
- **CompareDifferences** - A report of variable value differences

PROC COMPARE Code

```
ods exclude all;
ods output
    CompareDatasets          = __data0
    CompareVariables(nowarn) = __vars0
    CompareSummary           = __summ0
    CompareDifferences(nowarn) = __diff0;
proc compare base=Base compare=Compare listvar
    criterion=&criterion;
    %if &keys ne %then %do;
        id &keys;
    %end;
run;
ods select all;
```

CompareDatasets

Aspects to Consider

1. The procedure information is useful but not needed as part of the final CompareDatasets data.
2. The number of columns in the Data Set Summary section can change, depending on the data.
3. The Variables Summary section does not need to be divided into multiple pieces.

Strategy

1. Separate the procedure information into its own data set.
2. Take advantage of the Data Set Summary information being space delimited.
3. Copy the text of the Variables Summary section as is.

Visual Depiction

Separate {

type	batch
h	The COMPARE Procedure
h	Comparison of WORK.PROD_SPRINGS with WORK.VAL_SPRINGS1
h	(Method=EXACT)
h	
h	Data Set Summary
h	
h	Dataset Created Modified NVar NObs
d	
d	WORK.PROD_SPRINGS 28JUN25:21:51:13 28JUN25:21:51:13 7 1587
d	WORK.VAL_SPRINGS1 28JUN25:21:51:13 28JUN25:21:51:13 8 1587
d	
d	
h	Variables Summary
h	
d	Number of Variables in Common: 7.
d	Number of Variables in WORK.VAL_SPRINGS1 but not in WORK.PROD_SPRINGS: 1.
d	Number of Variables with Differing Attributes: 1.

Space Delimited

Leave As Is



Code for #1

```
data __head;
    set __data0 ;
    where &where2;
    length head1 $1000;

    retain head1 ;
    head1 = catx(' (*ESC*)n',head1,batch);
    head1 = tranwrd(head1,'WORK.BASE',upcase("&basedr..&baseds")) ;
    head1 = tranwrd(head1,'WORK.COMPARE',
        upcase("&compdr..&compds")) ;

    if _n_ = 3;
    keep head1;

run;
```

Code for #2

```
array info{*} var1-var&data_num;
if type = 'h' and indexw(batch, 'Dataset') then do;
*find starting place of label if it is present;
    %if &data_num = 6 %then %do;
        retain place;
        place = indexc(batch, 'L');
    %end;
end;
if section = 2 then do i=1 to &data_num;
    if i < 6 then info{i} = scan(batch, i, ' ');
    else info{i} = substr(batch, place);
end;
```

Code for #3

```
if section = 4 then var1 = strip(batch);
```

The section number is based on code not shown in the presentation.
See paper for the full code.

New CompareDatasets Data

	head2	var1	var2	var3	var4	var5
1	Data Set Summary	Dataset	Created	Modified	NVar	NObs
2	Data Set Summary					
3	Data Set Summary	WORK.BASE	14AUG25:15:29:58	14AUG25:15:29:58	7	1587
4	Data Set Summary	WORK.COMPARE	14AUG25:15:29:58	14AUG25:15:29:58	7	1587
5	Data Set Summary					
6	Data Set Summary					
7	Variables Summary	Number of Variables in Common: 7.				
8	Variables Summary	Number of Variables with Differing Attributes: 2.				

CompareVariables

Aspects to Consider

1. The number of columns in the Differing Attributes section can change, depending on the data.
2. The system option PAGESIZE determines how many data observations are grouped together.
3. The number of rows displayed together is affected by the amount of data.

Strategy

1. Count the number of sections.
2. Count how many pieces are in each section.
3. Determine the starting position of each piece within each section.

Visual Depiction

1 Section

type	batch							▲
d								
d								
h		Listing of Common Variables with Differing Attributes						
d								
h		Starting I						
h								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d								
d	</							

6 Pieces



35

46

60

66

74

82

Starting Positions



Code for #1

```
data __vars0a;  
  set __vars0;  
  where &where1;  
  retain section 0;  
  if index(batch, 'Listing of') then do;  
    section + 1;  
    call symputx('var_sect_lbl' ||  
                 strip(put(section,8.)), strip(batch));  
    delete;  
  end;  
run;
```

Code for #2

```
data _null_;  
    set __vars0a;  
    if type = 'h' and ^missing(batch);  
    n+1;  
    call symputx('var' || strip(put(n,8.)) ||  
        '_num',countw(batch));  
    call symputx('var_sect',n);  
  
run;  
  
VAR_SECT_LBL1=Listing of Common Variables with Differing Attributes  
VAR1_NUM=6  
VAR_SECT=1
```

Code for #3

```
data __positions_vars&i;  
  set __vars0a;  
  where type = 'h' and ^missing(batch) and section = &i;  
  
  array v1{*} vnum1-vnum&&var&i._num;  
  do i=1 to &&var&i._num;  
    if i=1 then v1{i} = anyfirst(batch);  
    else v1{i} = anyfirst(batch,v1{i-1}+len1);  
    len1 = length(scan(batch,i));  
  end;  
  
  keep section vnum;;  
run;
```

Divide Batch Variable

```
%do j = 1 %to &var_sect;
  %do jj = 1 %to &&var&j._num;
    if section = &j and type = 'd' then do;
      if &jj < &&var&j._num then info{&jj} =
        substr(batch,place{&jj},place{&jj+1}-place{&jj});
      else if &jj = &&var&j._num then info{&jj} =
        substr(batch,place{&jj});
    end;
  %end;
  if section = &j then head2 = "&&var_sect_lbl&j.";
  %if &j > 1 %then %do;
    %let jjj = %eval(&j-1);
    if section = &j and bold='Y' and
      head2 = "&&var_sect_lbl&jjj" then delete;
  %end;
%end;
```

New CompareVariables Data

	section	head2	var1	var2	var3	var4	var5	var6
1	0							
2	0							
3	1	Listing of Common Variables with Differing Attributes						
4	1	Listing of Common Variables with Differing Attributes	Variable	Dataset	Type	Length	Format	Label
5	1	Listing of Common Variables with Differing Attributes						
6	1	Listing of Common Variables with Differing Attributes	Area	WORK.BASE	Char	14		
7	1	Listing of Common Variables with Differing Attributes		WORK.COMPARE	Char	14		Geographical Area
8	1	Listing of Common Variables with Differing Attributes	Fahrenheit	WORK.BASE	Num	8		
9	1	Listing of Common Variables with Differing Attributes		WORK.COMPARE	Num	8	8.1	

CompareSummary

Aspects to Consider

1. When the comparison includes an ID statement, the values of the ID variables are included after the Compare observation number. When those values are longer than the line size they wrap to the next observation.
2. The Observation Summary section is best split into two sections, making a total of four sections.
3. The number of columns in the Variables with Unequal Values section can change, depending on the data.

Strategy

1. Check the number of rows to determine if wrapping occurred.
2. Assign section numbers.
3. Count how many pieces are in the Variables with Unequal Value section and determine the starting position of each piece within each section.

Visual Depiction

Check for wrapping

	type	batch
1	d	
2	d	
3	h	Observation Summary
4	h	
5	h	Observation Base Compare
6	d	
7	d	First Obs 1 1
8	d	First Unequal 14 14
9	d	Last Unequal 1400 1400
10	d	Last Obs 1587 1587
11	d	
12	d	Number of Observations in Common: 1587.
13	d	Total Number of Observations Read from WORK.BASE: 1587.
14	d	Total Number of Observations Read from WORK.COMPARE: 1587.
15	d	
16	d	Number of Observations with Some Compared Variables Unequal: 3.
17	d	Number of Observations with All Compared Variables Equal: 1584.

Separate into
two sections

Visual Depiction (continued)

Section 3

20	h	Values Comparison Summary			
21	h				
22	d	Number of Variables Compared with All Observations Equal: 5.			
23	d	Number of Variables Compared with Some Observations Unequal: 2.			
24	d	Total Number of Values which Compare Unequal: 3.			
25	d	Maximum Difference Criterion Value: 0.24818.			
26	d				
27	h	The COMPARE Procedure			
28	h	Comparison of WORK.BASE with WORK.COMPARE			
29	h	(Method=RELATIVE(0.0000222), Criterion=1.0E-09)			
30	h				
31	h	Variables with Unequal Values			
32	h	48	59	65	70 76
33	h	Variable	Type	Len	Ndif MaxCrit
34	d				
35	d	Name	CHAR	34	1
36	d	Celsius	NUM	8	2 0.248

5 Pieces
└─→

Starting Positions

← These rows are removed

Code for #1

```
data _null_;
    set __summ0a end=eof;
        retain firstn lastn ;
        if indexw(batch,'First Obs') then do;
            firstn = _n_;
            call symputx('firstobs',_n_);
        end;
        if indexw(batch,'First Unequal') then firstn = _n_ - firstn;
        if index(batch,'Last Unequal') then lastn = _n_;
        if index(batch,'Last Obs') then do;
            lastn = _n_ - lastn;
            call symputx('lastobs',_n_);
        end;
        if eof then do;
            if firstn > 1 or lastn > 1 then call symputx('multiobservation','Y');
            else call symputx('multiobservation','N');
        end;
run;
```

Code for #1 (continued)

```
data __multi;  
  set __summ0a(rename=(batch=orig));  
  length dummy batch $500;  
    if _n_ < &firstobs then do; batch = orig; output; end;  
    if &firstobs <= _n_ <= &lastobs + 1;  
    dummy = lag(orig);  
    if index(dummy,'First Obs') and ^index(orig,'First Unequal') then do;  
      batch = catt(dummy, orig); output; end;  
    if index(dummy,'First Unequal') and ^index(orig,'Last Unequal') then do;  
      batch = catt(dummy, orig); output; end;  
    else if index(dummy,'Last Unequal') and ^index(orig,'Last Obs') then do;  
      batch = catt(dummy, orig); output; end;  
    else if index(dummy,'Last Obs') and ^missing(orig) then do;  
      batch = catt(dummy, orig); output; end;  
  keep type batch;  
run;
```

Code for #2

```
data __summ0d;  
  set __summ0c;  
  where &where1;  
    retain section 0;  
    if type = 'h' and (index(batch, 'Summary') or  
index(batch, 'Unequal')) then do;  
      section + 1;  
      call symputx('sum_sect_lbl' ||  
strip(put(section, 8.)), strip(batch));  
      delete;  
    end;  
run;
```

Code for #3

```
*determine how many pieces are in each section;  
data _null_;  
    set __summ0d;  
    if type = 'h' and ^missing(batch);  
        batch = tranwrd(batch, 'Compare Label',  
                        'Compare_Label');  
    call symputx('sum' || strip(put(section, 8.)) ||  
                '_num', countw(batch));  
    call symputx('sum_sect', section);  
  
run;
```

Code for #3 (continued)

```
data __positions_summ&i;  
  set __summ0d;  
  where type = 'h' and ^missing(batch) and section = &i;  
  batch = tranwrd(batch, 'Compare Label', 'Compare_Label');  
  array v1{*} vnum1-vnum&&sum&i._num;  
  len1 = length(scan(batch,1));  
  do i=1 to &&sum&i._num;  
    if i=1 then v1{i} = anyfirst(batch);  
    else v1{i} = anyfirst(batch,v1{i-1}+len1);  
    len1 = length(scan(batch,i));  
  end;  
  keep section vnum;;  
run;
```

New CompareSummary Data

	section	head2	var1	var2	var3
1	0				
2	1				
3	1	Observation Summary	Observation	Base	Compare
4	1	Observation Summary			
5	1	Observation Summary	First Obs	1	1
6	1	Observation Summary	First Unequal	14	14
7	1	Observation Summary	Last Unequal	1400	1400
8	1	Observation Summary	Last Obs	1587	1587
9	1	Observation Summary			
10	2	Observation Number Summary	Number of Observations in Comr		
11	2	Observation Number Summary	Total Number of Observations Re		
12	2	Observation Number Summary	Total Number of Observations Re		
13	2	Observation Number Summary			
14	2	Observation Number Summary	Number of Observations with Sor		
15	2	Observation Number Summary	Number of Observations with All		
16	2	Observation Number Summary			
17	2	Observation Number Summary			
18	3	Values Comparison Summary	Number of Variables Compared v		
19	3	Values Comparison Summary	Number of Variables Compared v		
20	3	Values Comparison Summary	Total Number of Values which Co		
21	3	Values Comparison Summary	Maximum Difference Criterion Va		
22	3	Values Comparison Summary			

New CompareSummary Data (continued)

25	4	Variables with Unequal Values	Variable	Type	Len	Ndif
26	4	Variables with Unequal Values				
27	4	Variables with Unequal Values	Name	CHAR	34	1
28	4	Variables with Unequal Values	Celsius	NUM	8	2
29	4	Variables with Unequal Values				

CompareDifferences

Aspects to Consider

1. The printed output and the CompareDifferences Output data set include observations filled with underscores (_).
2. The header before the vertical bars changes if an ID statement is used.
3. The vertical bars (|) used for alignment are also captured in the data set, but the SCAN function does not consider a vertical bar as a word.
4. The pieces that are displayed in the Value Comparison Results for Variables differ between character variables and numeric variables.
5. Some headers contain multiple words delimited by spaces and other default delimiters (“Base Value” and “% Diff”), which complicates parsing.

Strategy

1. Maintain the underscores but only use nine in each new variable.
2. Use the observations with underscores and vertical bars to determine the starting positions, not the header observation.
3. Separately identify the position of the vertical bar.
4. Count the pieces that aren't vertical bars. Sort the starting positions.
5. Use assignment statements to change the headers with multiple words.

Visual Depiction

	type	batch
1	d	
2	d	
3	h	Value Comparison Results for Variables
4	h	
5	d	
6	d	Base Value Compare Value
7	h	Obs Name Name
8	d	+
9	d	
10	d	88 Hot Springs Near Icy Hot Springs Near Icy
11	d	
12	d	
13	d	Manually changed
14	d	
15	d	Base Compare Diff. % Diff
16	h	Obs Celsius Celsius Diff. % Diff
17	d	
18	d	
19	d	14 154.0000 120.0000 -34.0000 -22.0779
20	d	1400 130.0000 120.0000 -10.0000 -7.6923
21	d	

Leave as is

Count separately

Manually change

Use to count pieces

Use this row to determine starting positions

Code for #1

```
if countc(batch, '_') > 20 then type = 'x';  
...  
if type = 'x' then info{i} = repeat('_', 9);
```

Code for #2, #3, #4

```
data __positions_diff&i;
  set __diff0a;
  *using the underscore observations because proc compare has
  inconsistent spacing when it mixes character and numeric;
  where index(batch, '_') and index(batch, '||') and section = &i; #2
    array v1{*} vnum1-vnum&&dif&i._num;
    do i=1 to &&dif&i._num;
      if i=1 then v1{i} = anyfirst(batch);
      else v1{i} = anyfirst(batch, v1{i-1}+len1); #4
      len1 = length(scan(batch, i));
      if i = &&dif&i._num then v1{i} = indexc(batch, '|'); #3
    end;
    call sortn(of v1{*}); #4
  keep section vnum;;
run;
```

Code for #5

```
batch = tranwrd(batch, '% Diff', '%_Diff');  
batch = tranwrd(batch, 'Diff.', 'Diff_');  
batch = tranwrd(batch, 'Base Value', 'Base_Value');  
batch = tranwrd(batch, 'Compare Value', 'Compare_Value');
```

New CompareDifferences Data

var1	var2	var3	var4	var5	var6
_____	_____	_____	_____		
		Base Value	Compare Value		
Obs		Name	Name		
_____	_____	_____	_____		
88		Hot Springs Near Icy	Hot Springs Near Icy		
_____	_____	_____	_____		
_____	_____	_____	_____	_____	_____
		Base	Compare		
Obs		Celsius	Celsius	Diff.	% Diff
_____	_____	_____	_____	_____	_____
14		154.0000	120.0000	-34.0000	-22.0779
1400		130.0000	120.0000	-10.0000	-7.6923
_____	_____	_____	_____	_____	_____

Final Excel Report

PROC REPORT Code

```
proc report data=tab_sum noheader;
    column grp head1 head2 var1-var&maxall
           bold;
    define grp / order noprint;
    define head1 / order noprint;
    define head2 / order noprint order=data;
    define var1 / style(column)=[width=3.5in];
    define bold / noprint;
compute before head1;
    if grp = 1 then num = 1000;
    else num = 0;
    line head1 $varying. num;
endcomp;
```

```
compute before head2;
    line head2 $1000.;
endcomp;
compute after head2;
    line ' ';
endcomp;
compute bold;
    if bold = 'Y' then call define
        (_row_,'style','style=[font_weight=bold]');
endcomp;
```

CompareDatasets and CompareVariables

2	Data Set Summary					
3	Dataset	Created	Modified	NVar	NObs	
4	WORK.BASE	03JUL25:15:19:36	03JUL25:15:19:36	7	1587	
5	WORK.COMPARE	03JUL25:15:19:36	03JUL25:15:19:36	7	1587	
6						
7	Variables Summary					
8	Number of Variables in Common: 7.					
9	Number of Variables with Differing Attributes: 2.					
10						

11	Listing of Common Variables with Differing Attributes					
12	Variable	Dataset	Type	Length	Format	Label
13	Area	WORK.BASE	Char	14		
14		WORK.COMPARE	Char	14		Geographical Area
15	Fahrenheit	WORK.BASE	Num	8		
16		WORK.COMPARE	Num	8	8.1	
17						

CompareSummary

18	Observation Summary				
19	Observation	Base	Compare		
20	First Obs	1	1		
21	First Unequal	14	14		
22	Last Unequal	1400	1400		
23	Last Obs	1587	1587		
24					
25	Observation Number Summary				
26	Number of Observations in Common: 1587.				
27	Total Number of Observations Read from WORK.BASE: 1587.				
28	Total Number of Observations Read from WORK.COMPARE: 1587.				
29	Number of Observations with Some Compared Variables Unequal: 3.				
30	Number of Observations with All Compared Variables Equal: 1584.				
31					
32	Values Comparison Summary				
33	Number of Variables Compared with All Observations Equal: 5.				
34	Number of Variables Compared with Some Observations Unequal: 2.				
35	Total Number of Values which Compare Unequal: 3.				
36	Maximum Difference Criterion Value: 0.24818.				
37					
38	Variables with Unequal Values				
39	Variable	Type	Len	Ndif	MaxCrit
40	Name	CHAR	34	1	
41	Celsius	NUM	8	2	0.248
42					

CompareDifferences

43	Value Comparison Results for Variables					
44						
45			Base Value	Compare Value		
46	Obs		Name	Name		
47						
48	88		Hot Springs Near Icy	Hot Springs Near Icy		
49						
50						
51			Base	Compare		
52	Obs		Celsius	Celsius	Diff.	% Diff
53						
54	14		154	120	-34	-22.0779
55	1400		130	120	-10	-7.6923
56						
57						

Questions?