

Tips from the Trenches

Uncommon Solutions to Common Problems

Christopher Bost

BASUG/Boston Area SAS Users Group
November 16, 2011

Recent user questions

2

- How can I print long character values?
- How can I EXPORT formatted values to Excel?
- How can I flag inconsistencies across observations?

How can I print long character values?

3

For example: To print direct quotes

- PROC PRINT
 - ▣ Truncates values to fit LINESIZE
- Output to Rich Text Format [RTF]
 - ▣ ODS RTF command
 - ▣ Enterprise Guide option
- PROC REPORT
 - ▣ WIDTH=n FLOW

PROC PRINT: Default output

4

```
0
b
s a                                b

1 1 The nation's retailers on Thursday confirmed that they had a merry Christmas in 2009, a year aft
2 2 Despite extensive government intervention in the housing market, some policy makers at the Feder
3 3 Even as concerns rise about technology distractions for drivers, automakers are rapidly bringing

0
b
s c

1 X
2 Y
3 Z
```

**WARNING: Data too long for column "b";
truncated to 96 characters to fit.**

Output to Rich Text Format [RTF]

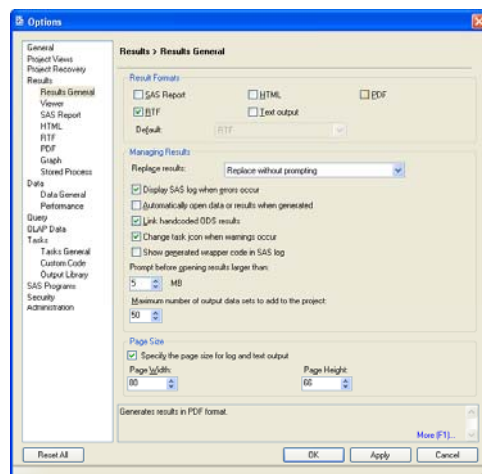
5

```
ods rtf file='c:\news.rtf';  
proc print data=news;  
run;  
ods rtf close;
```

Obs	a	b	c
1	1	The nation's retailers on Thursday confirmed that they had a merry Christmas in 2009, a year after weathering the worst holiday shopping season in decades.	X
2	2	Despite extensive government intervention in the housing market, some policy makers at the Federal Reserve are worried that even more might need to be done.	Y
3	3	Even as concerns rise about technology distractions for drivers, automakers are rapidly bringing PC features to the dashboard.	Z

Set Enterprise Guide RTF option

6



1. Open **T**ools menu
2. Select **O**ptions
3. Results General
4. Check RTF box

PROC REPORT: Syntax

7

```
proc report data=news headline headskip;  
column a b c ;  
define a / order;  
define b / display width=25 flow;  
define c / display;  
break after a / skip;  
run;
```

PROC REPORT: Output

8

	a b	c
1	The nation's retailers on Thursday confirmed that they had a merry Christmas in 2009, a year after weathering the worst holiday shopping season in decades.	X
2	Despite extensive government intervention in the housing market, some policy makers at the Federal Reserve are worried that even more might need to be done.	Y
3	Even as concerns rise about technology distractions for drivers, automakers are rapidly bringing PC features to the dashboard.	Z

How can I export formatted values to Excel?

9

For example: To send formatted values to a client

- PROC EXPORT to Excel
 - ▣ Exports “internal” values
 - ▣ Neither numeric nor character formats are used
- PROC EXPORT to CSV
 - ▣ CSV = Comma-Separated Values
 - ▣ Exports formatted values
 - ▣ Excel can open CSV files
 - Save as .XLS

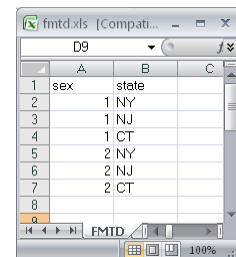
PROC EXPORT to Excel

10

```
proc print data=fmtd;  
run;
```

Obs	sex	state
1	1: Female	NY: New York
2	1: Female	NJ: New Jersey
3	1: Female	CT: Connecticut
4	2: Male	NY: New York
5	2: Male	NJ: New Jersey
6	2: Male	CT: Connecticut

```
proc export  
data=fmtd  
outfile='c:\fmtd.xls'  
dbms=excel;  
run;
```



PROC EXPORT to CSV

11

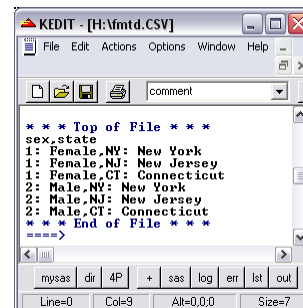
```
proc export  
data=fmtd  
outfile='c:\fmtd.csv'  
dbms=csv;  
run;
```

A screenshot of a Microsoft Excel spreadsheet. The title bar says "fmtd.csv - Micros...". The spreadsheet has two columns, A and B. Row 1 contains the headers "sex" and "state". Rows 2 through 7 contain data: "1: Female NY: New York", "1: Female NJ: New Jersey", "1: Female CT: Connecticut", "2: Male NY: New York", "2: Male NJ: New Jersey", and "2: Male CT: Connecticut". Row 8 is empty. The status bar at the bottom shows "fmtd", "100%", and a zoom slider.

What is a CSV file?

12

- ASCII file
 - open with any text editor
 - Excel knows how to **read**
- Comma-Separated Values
- First row contains column names
- Additional rows contain values
 - separated by commas



Save CSV file as Excel

13

- Save As
- Save as type:
 - ▣ Excel Workbook (*.xlsx)
 - ▣ Excel 97-2003 Workbook (*.xls)
- Note current type is CSV (Comma delimited) (*.csv)
 - ▣ .CSV file extension usually registered to Excel
 - ▣ Can change registration in Windows Explorer

Associating extensions with applications

14

A file extension can be associated with an application so that when a file with that extension is double-clicked in Windows Explorer it is opened with that application. Work with registered file extensions as follows:

1. Open Windows Explorer
2. Open the Tools menu and select Folder Options...
3. Click the "File Types" tab. [There might be a delay before registered files types load.]
4. Scroll to the file extension of interest; e.g., CSV. [Hint: Type the first letter (only) to jump to file extensions that start with that letter.]
5. Click the "Change..." button to change the application with which files with this extension should be opened. The "Open With" dialog box appears.
6. Scroll to the application of interest; e.g., Microsoft Office Excel.
7. Click the OK button.
8. Repeat as desired for other file extensions. If a file extension is not a registered file type, click the New button, enter the File Extension, click OK, and repeat the process.
9. Click the Close button when finished.

How can I flag inconsistencies?

15

For example: To check multiple records per student

- Values of demographic variables should not change
- Flag values that are not the same

Sample data

16

Obs	id	x	y	sex	age
1	1	11	22	F	10
2	2	33	44	M	11
3	2	33	44	M	11
4	3	55	66	F	12
5	3	55	66	F	12
6	3	55	66	F	13
7	4	77	88	M	14
8	4	77	88	F	14

PROC SQL: Syntax

17

```
proc sql;
create table flags as
select *,
       case when count(distinct age) > 1
            then '*'
            else ' '
       end as ageflag,
       case when count(distinct sex) > 1
            then '*'
            else ' '
       end as sexflag
from (select *, count(id) as Frequency
      from k12
      group by id,x,y,sex,age)
group by id
having count(id) > 1;
quit;
```

FROM clause is executed first

18

```
proc sql;
create table flags as
select *,
       case when count(distinct age) > 1
            then '*'
            else ' '
       end as ageflag,
       case when count(distinct sex) > 1
            then '*'
            else ' '
       end as sexflag
from (select *, count(id) as Frequency
      from k12
      group by id,x,y,sex,age)
group by id
having count(id) > 1;
quit;
```

A query in parentheses on a FROM clause is called an in-line view.

In-Line view produces a “virtual table”

19

id	x	y	sex	age	Frequency
1	11	22	F	10	1
2	33	44	M	11	2
3	55	66	F	12	2
3	55	66	F	13	1
4	77	88	F	14	1
4	77	88	M	14	1

GROUP BY and HAVING executed next

20

```
proc sql;
create table flags as
select *,
       case when count(distinct age) > 1
            then '*'
            else ' '
       end as ageflag,
       case when count(distinct sex) > 1
            then '*'
            else ' '
       end as sexflag
from (select *, count(id) as Frequency
      from k12
      group by id,x,y,sex,age)
group by id
having count(id) > 1;
quit;
```

GROUP BY and HAVING: Results

21

id	x	y	sex	age	Frequency
3	55	66	F	12	2
3	55	66	F	13	1
4	77	88	F	14	1
4	77	88	M	14	1

SELECT is executed last

22

```
proc sql;
create table flags as
select *,
      case when count(distinct age) > 1
        then '*'
        else ' '
      end as ageflag,
      case when count(distinct sex) > 1
        then '*'
        else ' '
      end as sexflag
from (select *, count(id) as Frequency
      from k12
      group by id,x,y,sex,age)
group by id
having count(id) > 1;
quit;
```

PROC PRINT with BY ID

23

id=3

Obs	x	y	sex	age	Frequency	ageflag	sexflag
1	55	66	F	12	2	*	
2	55	66	F	13	1	*	

id=4

Obs	x	y	sex	age	Frequency	ageflag	sexflag
3	77	88	F	14	1		*
4	77	88	M	14	1		*

Summary of SQL steps

24

- Use GROUP BY in an in-line view to create a row for each unique combination of specified variables
- Use GROUP BY and HAVING to subset rows where ID occurs more than once
- Select all columns and use CASE to create variables that flag values that are not constant

Recent user questions

25

- How can I create unique SAS data set names?
- How can I control the length of formatted values?
- How can I count unique values within an observation?

How can I create unique SAS data set names?

26

- Problem
 - ▣ Program creates SAS data set[s]
 - ▣ Program is run repetitively
 - ▣ Each output data set must be saved
 - ▣ Data set names must be unique to avoid overwriting
 - ▣ Do **not** want to have to edit program each time!
- Solution
 - ▣ Create macro variable with current date-time “stamp”
 - ▣ Use macro variable when creating SAS data set name

Create unique values with DATETIME

27

- DATETIME()
 - ▣ Returns **1578493853.3**
 - ▣ Number of seconds from January 1, 1960 at midnight
 - ▣ Not particularly useful as-is

- DATETIME() formatted with DATETIME20.
 - ▣ Returns **07JAN2010:14:30:53**
 - ▣ Useful as a unique “date-time stamp”
 - ▣ Need to change colons to underscores

Create date-time stamp with DATA step

28

```
data test;  
dt1 = datetime();  
dt2 = put(dt1,datetime20.);  
dt3 = translate(dt2,'_',':');  
put dt1= / dt2= / dt3= ;  
run;
```

produces:

```
dt1=1578493853.3  
dt2=07JAN2010:14:30:53  
dt3=07JAN2010_14_30_53
```

Problem: We want to **create a macro variable with %LET**, but there are no macro function equivalents of the DATA step functions PUT and TRANSLATE.

Create date-time stamp with %LET

29

```
%let dtstamp = %sysfunc(translate  
(%sysfunc(datetime()),datetime20.),'_',':');
```

- **%SYSFUNC** executes DATA step functions*
- First %SYSFUNC is for TRANSLATE function
- Second %SYSFUNC is for DATETIME function
 - ▣ Note optional format assigned to result

*Not all DATA step functions are available with %SYSFUNC [e.g., PUT and INPUT].

Using date-time stamp

30

```
data sasdata.test&dtstamp;
```

```
...
```

```
run;
```

creates data set

test07JAN2010_14_30_53

How can I control length of formatted values?

31

- Problem
 - ▣ Format labels some, but not all, values
 - ▣ Want unformatted values to print as-is
 - ▣ Some printed values are truncated
- Solution 1
 - ▣ (DEFAULT=*n*)
- Solution 2
 - ▣ OTHER=[*format.*]

Sample program

32

```
proc format fmtlib;
value $schoolf 'Brooklyn Latin School',
               'Brooklyn Technical High School' = 'SHSAT';
run;
```

↓
SHSAT = Specialized High
Schools Admissions Test

```
data temp;
input school $ 1 - 30 ;
format school $schoolf. ;
datalines;
Brooklyn College Academy
Brooklyn High School of the Arts
Brooklyn School for Global Studies
Brooklyn Technical High School
Brooklyn Latin School
;

proc print data=temp;
run;
```

} print original values
} print formatted values ("SHSAT")

Formatted results

33

Obs	school
1	Brook
2	Brook
3	Brook
4	SHSAT
5	SHSAT



FMTLIB output

34

```
-----  
|      FORMAT NAME: $SCHOOLF LENGTH:      5  NUMBER OF VALUES:      2  |  
|  MIN LENGTH:      1  MAX LENGTH:     40  DEFAULT LENGTH  5  FUZZ:      0  |  
|-----|  
| START          | END          | LABEL  (VER. V7|V8   01DEC2008:22:10:15) |  
|-----+-----+-----|  
| Brooklyn Latin S| Brooklyn Latin S| SHSAT                                     |  
| Brooklyn Technic| Brooklyn Technic| SHSAT                                     |  
|-----|
```

Note: The **default length** of a format is the length of the longest formatted value.

Solutions: DEFAULT and existing format

35

Solution 1

```
proc format;  
value $schoolf(default=40)  
    'Brooklyn Latin School',  
    'Brooklyn Technical High School' = 'SHSAT';  
run;
```

Solution 2

```
proc format;  
value $schoolf 'Brooklyn Latin School',  
    'Brooklyn Technical High School' = 'SHSAT'  
    OTHER = [$40.] ;  
run;
```

↑ ↑
square brackets
are required

Formatted results

36

Obs	school
1	Brooklyn College Academy
2	Brooklyn High School of the Arts
3	Brooklyn School for Global Studies
4	SHSAT
5	SHSAT

How can I count unique values within obs?

37

- Problem
 - ▣ Subjects have multiple SSNs from different sources
 - ▣ Count how many distinct values of SSN are used
- Solution 1
 - ▣ Restructure observations
 - ▣ Use COUNT DISTINCT in PROC SQL
 - ▣ Merge results back onto source data
- Solution 2
 - ▣ Single DATA step and the TRANWRD function

Sample data

38

Obs	id	ssn1	ssn2	ssn3	ssn4	ssn5
1	1	111	111	111	111	111
2	2	222	222	333	444	444
3	3	333	333	444	444	555
4	4	444	444		555	555

Solution 1

39

```
data one;
set many;
array ssns {5} ssn1 - ssn5 ;
do I = 1 to 5 ;
    ssn = ssns{I} ;
    output;
end;
keep id ssn ;
run;
```

Obs	id	ssn
1	1	111
2	1	111
3	1	111
4	1	111
5	1	111
6	2	222
7	2	222
8	2	333
9	2	444
10	2	444
11	3	333
12	3	333
13	3	444
14	3	444
15	3	555
16	4	444
17	4	444
18	4	
19	4	555
20	4	555

```
proc print data=one;
run;
```

Solution 1 [continued]

40

```
proc sql;
create table check as
select id, count(distinct ssn)
    as DistinctSSNs
from one
group by id;
quit;
```

Obs	id	Distinct SSNs
1	1	1
2	2	3
3	3	3
4	4	2

```
data merged;
merge many check;
by id;
run;
```

Obs	id	ssn1	ssn2	ssn3	ssn4	ssn5	Distinct SSNs
1	1	111	111	111	111	111	1
2	2	222	222	333	444	444	3
3	3	333	333	444	444	555	3
4	4	444	444		555	555	2

Solution 2

Single DATA step

41

```
data many2;
set many;
unique=0;
ssnx=catx(' ',of ssn1-ssn5);
array ssns {5} $ ssn1 - ssn5 ;
do I = 1 to 5 ;
    if ssn{I} ne ' ' and
        indexw(ssnx,ssns{I}) then do;
        unique=unique+1;
        ssnx = tranwrd(ssnx,ssns{I},' ');
    end;
end;
drop ssnx i ;
run;
```

Obs	id	ssn1	ssn2	ssn3	ssn4	ssn5	unique
1	1	111	111	111	111	111	1
2	2	222	222	333	444	444	3
3	3	333	333	444	444	555	3
4	4	444	444		555	555	2

Processing logic

42

Obs	id	ssn1	ssn2	ssn3	ssn4	ssn5
2	2	222	222	333	444	444

id=2 ssnx=222 222 333 444 444 unique=0

I=1 ssn1=222

ssnx=333 444 444 unique=1

I=2 ssn2=222

I=3 ssn3=333

ssnx=444 444 unique=2

I=4 ssn4=444

ssnx= unique=3

I=5 ssn5=444

Recent user questions

43

- How can I
 - ▣ determine the number of levels of a variable?
 - ▣ determine the number of combinations of vars?
 - ▣ run PROC CONTENTS on all data sets in folder?

Determine number of levels of variable

44

- Problem
 - ▣ Data set has hundreds [or thousands] of variables
 - ▣ Need to know how many levels the variables have
 - ▣ Do **not** want to run PROC FREQ on every variable!
- Solutions
 - ▣ COUNT DISTINCT in PROC SQL
 - ▣ NLEVELS option in PROC FREQ

Sample data set TEMP

45

obs	x	y
1	1	1
2	2	2
3	2	3
4	3	4
5	3	4
6	3	5
7	4	5
8	5	6
9	.	6

- How many unique values of X?
- How many unique values of Y?
- Do you count missing values?

SQL approach #1

46

```
proc sql;  
select count(x) as CountX  
from temp;  
quit;
```

*COUNT counts
nonmissing values.*

```
CountX  
-----  
      8
```

SQL approach #2

47

```
proc sql;
select count(distinct x)
      as CountDistinctX
from temp;
quit;
```

COUNT DISTINCT
counts *unique*
nonmissing values.

```
      Count
      DistinctX
-----
              5
```

FREQ approach #1

48

```
proc freq data=temp nlevels;
tables x;
run;
```

NLEVELS produces
a separate summary
of the N of levels

The FREQ Procedure

Number of Variable Levels

Variable	Levels	Missing Levels	Nonmissing Levels
x	6	1	5

x	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	1	12.50	1	12.50
2	2	25.00	3	37.50
3	3	37.50	6	75.00
4	1	12.50	7	87.50
5	1	12.50	8	100.00

FREQ approach #2

49

```
proc freq data=temp nlevels;  
tables x/noprint;  
run;
```

Turn off printing of
frequency table

The FREQ Procedure

Number of Variable Levels			
Variable	Levels	Missing Levels	Nonmissing Levels

x	6	1	5

FREQ approach #3

50

```
proc freq data=temp nlevels;  
tables x y/noprint;  
run;
```

NLEVELS produces
a separate summary
for multiple vars

The FREQ Procedure

Number of Variable Levels			
Variable	Levels	Missing Levels	Nonmissing Levels

x	6	1	5
y	6	0	6

FREQ approach #4

51

```
ods output nlevels=counts;  
proc freq data=temp nlevels;  
tables x y/noprint;  
run;
```

ODS OUTPUT saves
NLEVELS results to
specified data set

```
proc print data=counts;  
/* where nlevels > 10; */  
run;
```

Obs	Table Var	NLevels	NMiss Levels	NNon Miss Levels
1	x	6	1	5
2	y	6	0	6

How can I determine the number of combinations of two variables?

52

□ Problem

- Need to know how many unique combinations of values occur in the data
- Do **not** want to run PROC FREQ on every pair!

□ Solutions

- PROC FREQ OUT= data set
- COUNT DISTINCT in PROC SQL

Sample data set TEMP2

53

Obs	x	y
1	1	1
2	2	1
3	2	1
4	3	1
5	3	2
6	3	2
7	4	1
8	4	1
9	4	.

- How many unique combinations of X and Y?
- Do you count missing values?

FREQ approach #1

54

```
proc freq data=temp2;
tables x*y;
run;
```

Table of x by y

x		y		
		1	2	Total
Frequency				
Percent				
Row Pct				
Col Pct				

1	1	1	0	1
		12.50	0.00	12.50
		100.00	0.00	
		16.67	0.00	

2	2	0	2	2
		25.00	0.00	25.00
		100.00	0.00	
		33.33	0.00	

3	1	2	3	3
		12.50	25.00	37.50
		33.33	66.67	
		16.67	100.00	

4	2	0	2	2
		25.00	0.00	25.00
		100.00	0.00	
		33.33	0.00	

Total	6	2	8	
	75.00	25.00	100.00	

Frequency Missing = 1

Traditional crosstab contains cells with Frequency of 0

FREQ approach #2

55

```
proc freq data=temp2 noprint;
table x*y/out=xtab;
run;
```

OUT= saves results
to specified data set

```
proc print data=xtab;
run;
```

Obs	x	y	COUNT	PERCENT
1	1	1	1	12.5
2	2	1	2	25.0
3	3	1	1	12.5
4	3	2	2	25.0
5	4	.	1	.
6	4	1	2	25.0

FREQ approach #3

56

```
proc freq data=temp2;
table x*y/list;
run;
```

/LIST produces
combinations in
tabular format

The FREQ Procedure

x	y	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	1	1	12.50	1	12.50
2	1	2	25.00	3	37.50
3	1	1	12.50	4	50.00
3	2	2	25.00	6	75.00
4	1	2	25.00	8	100.00

Frequency Missing = 1

FREQ approach #3

57

```
proc freq data=temp2;
table x*y/list missing;
run;
```

MISSING includes
missing values in
table

The FREQ Procedure

x	y	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	1	1	11.11	1	11.11
2	1	2	22.22	3	33.33
3	1	1	11.11	4	44.44
3	2	2	22.22	6	66.67
4	.	1	11.11	7	77.78
4	1	2	22.22	9	100.00

SQL approach

58

```
proc sql;
select
count(distinct catx(' ',x,y))
as CountDistinctX_Y
from temp2;
quit;
```

CATX function starts
with delimiter to use,
followed by two or
more arguments

```
Count
DistinctX_Y
-----
6
```

How can I run PROC CONTENTS on all SAS data sets in a folder?

59

- Problem
 - ▣ Need to run PROC CONTENTS on multiple data sets
 - ▣ Do **not** want to run PROC CONTENTS on each one!
- Solutions
 - ▣ Macro and list of data set names
 - ▣ DATA=libref._ALL_ option in PROC CONTENTS

Macro and list of data set names

60

```
%macro cont(dslist=);  
%let nds=%sysfunc(countw(&dslist));  
%do i=1 %to &nds;  
  proc contents  
    data=%scan(&dslist,&i,' ');  
  run;  
%end;  
%mend cont;  
  
%cont(dslist=dsn1 dsn2 dsn3)
```

Count the number of data set names and then iterate through the list with the %SCAN function.

DATA=libref._ALL_ option in PROC CONTENTS

61

```
libname libref 'path';  
proc contents data=libref._ALL_;  
run;
```

ALL will run PROC CONTENTS on all SAS data sets in the specified folder, in alphabetical order

Questions?

62

Christopher Bost
Director, Research Technology Unit
MDRC
16 East 34th Street
New York, NY 10016
(212) 340-8613
christopher.bost@mdrc.org