www.OASUS.ca

SAS Tips and Tricks: "How Not to be a SAS Dinosaur"

Shawna Brown
Socio-Economic Analysis and Modeling Division
Statistics Canada

"Come out of the desert of ignorance to the OASUS of knowledge"

www.OASUS.ca

FREQUENCY REPORTS BY NUMBER OF OCCURRENCES

You want to generate a report of frequency counts, with the most common values listed at the top of the report. The FREQ procedure generates the counts, but the report shows the category values in numeric or alphabetical order.

```
proc freq data=testfreq;
    tables edlev / noprint out=freqcounts;
run;
proc sort data=freqcounts; by descending count;
run;
proc print data=freqcounts(obs=10); var edlev Count
  Percent;
run;
```

www.OASUS.ca

FREQUENCY REPORTS BY NUMBER OF OCCURRENCES

Distribution of Highest Level of Education

Obs	edlev	COUNT	PERCENT
1	Non-univ. postsecndry cert.	6164	25.1756
2	Not Applicable	4858	19.8415
3	Graduated high school	3282	13.4047
4	Bachelor's degree	2106	8.6015
5	Some non-univ. postsecndry (no cert.)	1724	7.0413
6	9-10 yrs of elem/secndry schol	1630	6.6574
7	11-13 yrs of elem/secndry schol(didn't grad)	1280	5.2279
8	Some univ. (no certificate)	1272	5.1952
9	5-8 years of elementary	870	3.5533
10	Univ cert above Bch/Mast,law degr,med/dentst/ vetnry/optom degr/Doct(PhD)	860	3.5125

FREQUENCY REPORTS BY NUMBER OF OCCURRENCES

Specify the ORDER=FREQ option of PROC FREQ.

```
proc freq data=testfreq order=freq;
  tables edlev / nocum;
run;
```

Distribution of Highest Level of Education

The FREQ Procedure

[Highest] Educational Level

EDLEV	Frequency	Percent
Non-univ. postsecndry cert.	6164	25.18
Not Applicable	4858	19.84
Graduated high school	3282	13.40
Bachelor's degree	2106	8.60
Some non-univ. postsecndry (no cert.)	1724	7.04
9-10 yrs of elem/secndry schol	1630	6.65
11-13 yrs of elem/secndry schol(didn't grad)	1280	5.23
Some univ. (no certificate)	1272	5.20
5-8 years of elementary	870	3.55
Univ cert above Bch/Mast,law degr,med/dentst/	860	3.51

Ottawa Area SAS Users Society

www.OASUS.ca

REORDER VARIABLES IN A DATASET

You may want to view a dataset with your key variables at the beginning of the dataset.

You may want to export a SAS data set to an Excel worksheet. The order of the columns in the exported worksheet must be different from the order of the variables in the original dataset.

REORDER VARIABLES IN A DATASET

```
ods select Position(persist);
proc contents data=in1.mydata2 varnum; title "Original
   Variable Order";
run;

data neworder;
    length hhseq inseq wgt 4 prov 8 age 3 sex marst
   edlev ictot 8;
set in1.mydata2;
run;

proc contents data=neworder varnum; title "Revised
   Variable Order";
run;
```

Ottawa Area SAS Users Society

www.OASUS.ca

REORDER VARIABLES IN A DATASET

Original Variable Order
The CONTENTS Procedure
--Variables Ordered by Position--

#	Variable	Type	Len	Format
1	SEX	Num	8	
2	MARST	Num	8	MRST.
3	PROV	Num	8	PROV.
4	EDLEV	Num	8	EDLV.
5	ICTOT	Num	6	9.
6	WGTHH	Num	4	8.
7	HHSEQ	Num	4	5.
8	INSEQ	Num	4	6.
9	AGE	Num	3	2.

Revised Variable Order
The CONTENTS Procedure
--Variables Ordered by Position-

#	Variable	Type	Len	Format
1	HHSEQ	Num	4	5.
2	INSEQ	Num	4	6.
3	WGT	Num	4	8.
4	PROV	Num	8	PROV.
5	AGE	Num	3	2.
6	SEX	Num	8	
7	MARST	Num	8	MRST.
8	EDLEV	Num	8	EDLV.
9	ICTOT	Num	8	9.

Ottawa Area SAS Users Society

www.OASUS.ca

REORDER VARIABLES IN A DATASET

Declare the desired order of the variable names in a RETAIN statement.

```
ods select Position(persist);
data neworder;
   retain hhseq inseq wgt prov age sex marst edlev
   ictot;
set in1.mydata2;
run;

proc contents data=neworder varnum; title "Revised
   Variable Order";
run;
```

*Output is the same as in the original approach.

Ottawa Area SAS Users Society

www.OASUS.ca

FILTERING OBSERVATIONS

You need to create a simple listing report limited to selected observations in an existing SAS dataset. The filtering can be accomplished using simple comparison criteria.

```
data t1snaf03;
    set t1sna03;
   if (status_code = 9); if (rprov < 10) or (tprov <
   10);
run;
proc print data=t1snaf03; run;
```

```
NOTE: There were 24225689 observations read from the data set IN1.T1SNAF03.
    NOTE: The data set WORK.T1SNA03 has 23634469 observations and 32 variables.
    NOTE: DATA statement used (Total process time):
          real time           44:03.14
          cpu time            14:26.71
```

FILTERING OBSERVATIONS

Apply the filter through a WHERE statement in the PROC PRINT step.

```
NOTE: There were 24225689 observations read from the data set IN1.T1SNA03.
NOTE: The data set WORK.T1SNA03 has 24225689 observations and 32 variables.
NOTE: DATA statement used (Total process time):
      real time           38:50.53
      cpu time            14:58.19
```

```
proc print data=t1sna03;
  where (status_code = 9) and ((rprov < 10) or (tprov < 10));
run;
```

www.OASUS.ca

FILTERING DATA: MISSING VALUES

You need to write a general macro program to filter missing data using a WHERE clause.

The WHERE statement requires type compatibility between operands, so a numeric variable uses the variable=. syntax while a character variable uses the variable=' ' syntax

FILTERING DATA: MISSING VALUES

```
%macro WhereMissing(var,type=N);
  %if &type=N %then %let missingvalue=.;
  %else %let missingvalue=' ';
  &var = &missingvalue
%mend WhereMissing;
proc print data=testmiss;
  title "Missing Sex and Province values";
  where %WhereMissing(sex,type=C)
    and %WhereMissing(prov,type=N)
  ;
run;
```

```
NOTE: There were 29 observations read from the data set
   WORK.TESTMISS.
      WHERE (sex=' ') and (prov=.);
```

www.OASUS.ca

FILTERING DATA: MISSING VALUES

Missing Sex and Province values

Obs	ICTOT	HHSEQ	AGE	sex	prov
12665	60000	30503	28		.
12666	70000	30504	44		.
12670	40500	30508	32		.

FILTERING DATA: MISSING VALUES

Use the IS MISSING operator in the WHERE expression.

```
%macro WhereMissing(var);
  &var is missing
%mend WhereMissing;
proc print data=testmiss;
  title "Missing Sex and Province values";
  where %WhereMissing(sex)
    and %WhereMissing(prov)
  ;
run;
```

```
NOTE: There were 29 observations read from the data set
   WORK.TESTMISS.
      WHERE (sex is null) and (prov is null);
```

*Output is the same as in the original approach.

www.OASUS.ca

VARIABLES WITH SIMILAR NAMES

You must process multiple related variables whose names start with the same character(s).

A list of variable names may be referenced include statements such as VAR in PROC PRINT, data set options such as KEEP= and DROP=, and functions such as SUM and MEAN.

Ottawa Area SAS Users Society

www.OASUS.ca

VARIABLES WITH SIMILAR NAMES

```
data income;
  set in1.mydatavar;
  where marst=4;
  totwork=workemp+worksenf;
  Work07=workemp+worksenf; Tot07=totwork+totinv+totpens;
run;
```

Use a name prefix list to reference multiple variable names.

```
data income;
     set mydata2;
     where marst=4;
      totwork=workemp+worksenf
     Work07=sum(of work:); Tot07=sum(of tot:);
   run;

proc print data=income (obs=5) var work: tot: run;
```

*Output is the same as in the original approach

www.OASUS.ca

FORMATS: NEGATIVES AND PERCENTAGES

You want to display negative numeric values with surrounding parentheses. You also want to display numeric values between 0.00 and 1.00 as percentages by adding the percent sign (%) as a suffix to the number.

Ottawa Area SAS Users Society

www.OASUS.ca

FORMATS: NEGATIVES AND PERCENTAGES

```
proc format;
  picture pcntage
          . = 'Missing'
       other = '00009.9%' (mult=1000);
  picture negative
          . = 'Missing'
      low-<0 = '000,000)' (prefix='(')
      0-high = '000,000 ' (prefix=' ');
run;
data income; set in1.mydata2;
     EmpasTot=0;
     if iemp ne 0 then EmpasTot=iemp/ictot;
     OthTot=1-EmpasTot;
     run;

proc print data=income(obs=5); format isenf negative8.; run;
proc print data=income(obs=5); format OthTot EmpasTot pcntage7.; run;
```

FORMATS: NEGATIVES AND PERCENTAGES

Self-Employment Income

Obs	prov	age	ictot	isenf
3	Alberta	80	37395	9,786
4	British Columbia	35	55132	(1,844)
71	British Columbia	28	21895	21,895
79	Alberta	25	36659	(611)
141	British Columbia	27	15019	(480)

Distribution between Employment and All Other Income

Obs	prov	age	iemp	ictot	Tot	OthTot
1	Alberta	46	12760	16368	77.9%	22.0%
3	Alberta	80	0	37395	0.0%	100.0%
5	B.C.	30	59020	59619	98.9%	1.0%
6	Alberta	48	28703	28703	100.0%	0.0%
10	B.C.	66	93	43633	0.2%	99.7%

Ottawa Area SAS Users Society

www.OASUS.ca

FORMATS: NEGATIVES AND PERCENTAGES

Specify the built-in NEGPARENw.d format, or the built-in PERCENTw.d format.

```
proc print data=income(obs=5);
  format isenf negparen8.;
  run;
proc print data=income(obs=5);
  format OthTot EmpasTot percent7.;
  run;
```

*Output is the same as in the original approach.

Ottawa Area SAS Users Society

www.OASUS.ca

DUPLICATE OBSERVATIONS: REMOVE COMPLETE DUPLICATES

You need to sort a dataset and remove duplicate observations at the same time.

The NODUPRECS option in PROC SORT is an easy way to request the removal of the duplicates. However, to guarantee proper removal of all duplicates, you must list every variable in the data set on the BY statement.

DUPLICATE OBSERVATIONS: REMOVE COMPLETE DUPLICATES

```
proc sort data=dupl out=unique noduprecs;
  by prov sex age hhseq edlev marst iemp ictot;
run;
```

```
NOTE: 82 duplicate observations were deleted.
NOTE: There were 24484 observations read from the data
   set WORK.DUPL.
NOTE: The data set WORK.UNIQUE has 24402 observations
   and 8 variables.
```

DUPLICATE OBSERVATIONS: REMOVE COMPLETE DUPLICATES

Reference all of the variables in the data set through the _ALL_ keyword.

```
proc sort data=dupl out=unique noduprecs;
  by prov sex _ALL_;
run;
```

```
NOTE: Duplicate BY variable(s) specified. Duplicates will be
   ignored.
NOTE: 82 duplicate observations were deleted.
NOTE: There were 24484 observations read from the data set
   WORK.DUPL.
NOTE: The data set WORK.UNIQUE has 24402 observations and 8
   variables.
```

Ottawa Area SAS Users Society

www.OASUS.ca

MICROSOFT ACCESS: WRITE SELECTED VARIABLES TO A TABLE

You need to export a SAS dataset to a Microsoft Access table.

You must subset the SAS data, both in terms of variables and observations, prior to performing the export. You also must assign a different name for some of the variables.

Ottawa Area SAS Users Society

www.OASUS.ca

MICROSOFT ACCESS: WRITE SELECTED VARIABLES TO A TABLE

```
data alberta;
  set in1.mydata2;
  where prov=8;
  keep prov hhseq inseq marst age;
  rename marst=MaritalStat;
run;

proc export data=alberta
            outtable=Alberta dbms=access replace;
  database="C:\shawna\alberta.mdb";
run;
```

Ottawa Area SAS Users Society

www.OASUS.ca

MICROSOFT ACCESS: WRITE SELECTED VARIABLES TO A TABLE

Apply data set options within the PROC EXPORT step.

```
proc export data=in1.mydata2
                    (where=(prov=8)
                     keep=prov hhseq inseq marst age
                     rename=(marst=MaritalStat))
              outtable=Alberta dbms=access replace;
  database="C:\shawna\alberta.mdb";
run;
```

Ottawa Area SAS Users Society

www.OASUS.ca

COMBINING EXCEL DATA WITH SAS DATA

You have two Excel worksheets that must be merged with a SAS dataset.

After the data is combined, the result must be converted back into a new Excel worksheet.

Ottawa Area SAS Users Society

www.OASUS.ca

COMBINING EXCEL DATA WITH SAS DATA

```
proc import datafile="C:\shawna\data\employ2.xls"
            dbms=Excel out=work.employ;
  Sheet="employ$";
run;
proc import datafile="C:\shawna\data\employ2.xls"
            dbms=Excel out=work.semploy;
  Sheet="semploy$";
run;

data inv; set invo; run;

…MERGE…

proc export data=WorkWithInvest dbms=Excel replace
            outfile="C:\shawna\excel\combined.xls" ;
  Sheet="IncomeExp";
run;
```

www.OASUS.ca

COMBINING EXCEL DATA WITH SAS DATA

Employment Income combined with Investment Income

inseq	prov	iemp	isenf	iinv
4	B.C.	58136	-1844	-1555
5	B.C.	59020	0	559
16	Ontario	9015	0	19141
42	PEI	21805	0	8967
120	Quebec	376	0	-548

COMBINING EXCEL DATA WITH SAS DATA

Use the Excel LIBNAME engine

```
libname myxls "C:\shawna\data\employ2.xls";

data semp; set myxls.'semploy$'n; run;
data emp; set myxls.'employ$'n; run;
data inv; set invo; run;
```

…MERGE…

```
proc export data=WorkWithInvest dbms=Excel replace
            outfile="C:\shawna\excel\combinednew.xls" ;
  Sheet="IncomeExp";
run;
```

*Output is the same as in the original approach

Ottawa Area SAS Users Society

www.OASUS.ca

Questions / Comments

Statistics Canada | Statistique Canada

Shawna Brown

Economist Researcher
Socio-Economic Analysis and Modeling Division

100 Tunney's Pasture Driveway
Ottawa, Ontario, Canada K1A 0T6
(613) 951-3607 Fax (613) 951-3959

Shawna.Brown@statcan.ca

Canada