

# SAS V8 Treasures

Charles Patridge

Email: [Charles\\_S\\_Patridge@prodigy.net](mailto:Charles_S_Patridge@prodigy.net)

Web: <http://www.sconsig.com>

1. An Example of using Multilabel Option in Proc Format SAS V8.2 to generate multi-level summaries/statistics
2. Get New Quantile Statistics in Proc Summary/Mean
3. Create Complete Tables when no records exist for certain Classes
4. Easily Create Top/Bottom ## of a list

```
/** Notice the Repeated USES of Values on left side of Equal Signs  
**/
```

```
Proc Format;
```

```
Value $REGNFMT (multilabel) /** NEW OPTION for Proc Format **/
```

```
CT, MA, RI, VT, NH, ME = "1 New England" /** New England **/
```

```
CT = "1 CT" MA = "1 MA" RI = "1 RI" VT = "1 VT" NH = "1 NH"
```

```
ME = "1 ME"
```

```
NY, NJ, PA, DE, MD, DC, VA, WV = "2 Mid Atlantic"/**MidAtlantic**/
```

```
NY = "2 NY" NJ = "2 NJ" PA = "2 PA" DE = "2 DE" MD = "2 MD"
```

```
DC = "2 DC" VA = "2 VA" WV = "2 WV"
```

```
SC, NC, GA, AL, FL = "3 South East" /** South East **/
```

```
SC = "3 SC" NC = "3 NC" GA = "3 GA" AL = "3 AL" FL = "3 FL"
```

```
OH, TN, KY, IL, IN, MI, WI = "4 Central North"/**CentralNorth **/
```

```
OH = "4 OH" TN = "4 TN" KY = "4 KY" IL = "4 IL" IN = "4 IN"
```

```
MI = "4 MI" WI = "4 WI"
```

```
LA, MS, MO, TX, AR, OK, KS = "5 Central South"/**CentralSouth **/
```

```
LA = "5 LA" MS = "5 MS" MO = "5 MO" TX = "5 TX" AR = "5 AR"
```

```
OK = "5 OK" KS = "5 KS"
```

```
IA, NE, SD, ND, MN = "6 Mid West" /*** Mid West      ***/  
IA = "6 IA"      NE = "6 NE"      SD = "6 SD"      ND = "6 ND"      MN = "6 MN"
```

```
WY, MT, ID, OR, WA = "7 North West" /*** North West  ***/  
WY = "7 WY"      MT = "7 MT"      ID = "7 ID"      OR = "7 OR"      WA = "7 WA"
```

```
CA, NV, UT, CO, NM, AZ = "8 South West" /*** South West ***/  
CA = "8 CA"      NV = "8 NV"      UT = "8 UT"      CO = "8 CO"      NM = "8 NM"  
AZ = "8 AZ"
```

```
/**/ Total USA      ***/  
CT, MA, RI, VT, NH, ME, NY, NJ, PA, DE, MD, DC, VA, WV, SC, NC, GA,  
AL, FL, OH, TN, KY, IL, IN, MI, WI, LA, MS, MO, TX, AR, OK, KS, IA,  
NE, SD, ND, MN, WY, MT, ID, OR, WA, CA, NV, UT, CO, NM, AZ  
= "0 Total USA"  
;
```

```
run;
```

```

options ls= 132 date center number pageno=1;

/**/ Will use the SAS Salary/Rate Survey Data to demonstrate /**/
/**/ Select only Full Time Employees /**/

Data Select;
  set sasuser.SASALARY;
  if "FULL TIME EMPLOYEE" = upcase( empltype );
  state = zipstate ( substr( left( zipcode ), 1, 5 ) );

region = Trim( left(put( state, $regfmt. )));

/**/ Now select only those states in the USA /**/
if substr(region,1,1) in ('1' '2' '3' '4' '5' '6' '7' '8');
***if substr(region,1,1) in ('0');

```

**NOTE:**

need to select on first assignments given in Proc Format when using MLF.

I could have used

```

if substr(region,1,1) in ('0');
if "0 Total USA" was given first in Proc Format
run;

```

```

/** Calculate Statistics using Proc Summary */
Proc Summary data=Select nway missing completetypes;
class state/MLF preloadfmt;
/**
NOTICE new option COMPLETETYPES
NOTICE new option MLF (Multilabel Format)
STATE grouped multiple times based upon the MLF and format $REGNFMT
NOTICE new option PRELOADFMT for creating entries that don't exist
**/
var salary;
output out=stats (drop=_type_ _freq_)
      n (salary) = count
      min (salary) = minimum
      p25 (salary) = q25          /** New Quantile Statistics */
      mean(salary) = average
      p50 (salary) = q50          /** New Quantile Statistics */
      p75 (salary) = q75          /** New Quantile Statistics */
      max (salary) = maximum;
format state $regnfmt. ;
run;
/** Print Report */
Title1 "Full Time SAS Employees - USA";
Title2 "Minimum, Average and Maximum SAS Salaries";
Proc Print data=stats ;
By state; Id state; Var _numeric_;
Format _numeric_ dollar10.0 count best.;
run;

```

Full Time SAS Employees - USA  
Minimum, Average and Maximum SAS Salaries

state	count	minimum	q25	average	q50	q75	maximum
0 Total USA	438	\$7,585	\$50,000	\$62,525	\$60,000	\$72,800	\$120,000
1 New England	44	\$7,585	\$50,000	\$66,070	\$64,000	\$81,000	\$112,000
1 CT	17	\$48,000	\$53,500	\$67,359	\$65,000	\$80,000	\$107,000
1 MA	25	\$7,585	\$50,000	\$66,399	\$68,000	\$82,000	\$112,000
1 ME	0	.	.	.	.	.	.
1 NH	1	\$61,000	\$61,000	\$61,000	\$61,000	\$61,000	\$61,000
1 RI	0	.	.	.	.	.	.
1 VT	1	\$41,000	\$41,000	\$41,000	\$41,000	\$41,000	\$41,000
2 Mid Atlantic	180	\$30,000	\$50,000	\$63,348	\$60,000	\$72,425	\$120,000
2 DC	16	\$30,000	\$56,500	\$70,504	\$69,534	\$83,250	\$115,000
2 DE	1	\$72,000	\$72,000	\$72,000	\$72,000	\$72,000	\$72,000
2 MD	36	\$35,000	\$45,000	\$56,042	\$51,500	\$61,200	\$106,000
2 NJ	38	\$39,000	\$52,000	\$61,663	\$60,000	\$68,000	\$100,000
2 NY	25	\$38,000	\$55,000	\$73,881	\$67,000	\$100,000	\$120,000
2 PA	40	\$41,000	\$53,500	\$63,179	\$61,700	\$72,525	\$106,000
2 VA	21	\$43,620	\$50,000	\$64,201	\$63,000	\$75,000	\$90,000
2 WV	3	\$34,500	\$34,500	\$39,833	\$40,000	\$45,000	\$45,000
3 South East	43	\$21,000	\$45,000	\$57,683	\$57,000	\$68,000	\$110,000
3 AL	2	\$39,800	\$39,800	\$47,400	\$47,400	\$55,000	\$55,000
3 FL	13	\$21,000	\$44,000	\$52,885	\$52,000	\$65,000	\$77,000
3 GA	7	\$36,000	\$60,000	\$64,760	\$70,000	\$75,000	\$80,000
3 NC	20	\$39,000	\$47,625	\$60,938	\$58,000	\$69,500	\$110,000
3 SC	1	\$25,989	\$25,989	\$25,989	\$25,989	\$25,989	\$25,989

Also, take a look at the various summaries that were generated with the use of Proc Format and Proc Summary (with MLF and PRELOADFMT option).

**COMPLETETYPES** - creates all possible combinations of class variables even if the combination does not occur in the input data set.

**Interaction:** The PRELOADFMT option in the CLASS statement ensures that PROC SUMMARY outputs all user-defined format ranges or values for the combinations of class variables, even when a frequency is zero.

take notice of where count = 0, entries that do not exist in data

```

%let top = 10;
proc means data=select noprint;
var salary;
output out=toplist
sum= mean=
idgroup( max(salary) out[&top] (salary state)=)/autolabel autoname;
run;
data toplist(keep=salary state);
set toplist;
array sal(*) salary_1--salary_&top;
array st (*) state_1--state_&top;
do i = 1 to dim(sal);
state = st(i); salary=sal(i); output;
end;
run;

title3 "Top &top Salaries";
proc print data=toplist; run;

```

Top 10 Salaries		
Obs	state	salary
1	NY	120000
2	NY	120000
3	NY	115000
4	DC	115000
5	MA	112000
6	NC	110000
7	CA	107500
8	MA	107000
9	CT	107000
10	PA	106000