Stefanie Schussler,

1. Given a dataset “work.claims” containing variables:

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Numeric/**  **Character** | **Length** | **Format** |
| Recip\_id | C | 9 | $9. |
| Prov\_id | C | 9 | $9. |
| Begin\_dos | N | 8 | mmddyy10. |
| End\_dos | N | 8 | mmddyy10. |
| Dx1 | C | 6 | $6. |
| Dx2 | C | 6 | $6. |
| Dx3 | C | 6 | $6. |
| Dx4 | C | 6 | $6. |
| Dx5 | C | 6 | $6. |

1. Write the code used to calculate the number of claims each provider had in 2000 (begin\_dos in 2000).

proc sql;

create table clm\_cnt as

select prov\_id, distinct(recip\_id) as clm\_cnt

from work.claims

where begin\_dos between ‘01JAN2000’d a ‘31DEC2000’d;

quit;

1. Write the code used to calculate the number of claims that had a diagnosis code between 290.xx and 319.xx coded for each provider in 2000 (begin\_dos in 2000).

proc sql;

create table clm\_cnt as

select prov\_id, distinct(recip\_id) as clm\_cnt

from work.claims

where begin\_dos between ‘01JAN2000’d a ‘31DEC2000’d

and (

substr(left(trim(dx1))1,1,3) ge ‘290’ and and substr(left(trim(dx1))1,1,3) le ‘319’

or substr(left(trim(dx2))1,1,3) ge ‘290’ and and substr(left(trim(dx2))1,1,3) le ‘319’

or substr(left(trim(dx3))1,1,3) ge ‘290’ and and substr(left(trim(dx3))1,1,3) le ‘319’

or substr(left(trim(dx4))1,1,3) ge ‘290’ and and substr(left(trim(dx4))1,1,3) le ‘319’

or substr(left(trim(dx5))1,1,3) ge ‘290’ and and substr(left(trim(dx5))1,1,3) le ‘319’

)

; quit;

1. Write the code used to calculate the number of times a diagnosis code between 290.xx and 319.xx was coded for each provider in 2000 (begin\_dos in 2000).

(if you can, please use a macro or an array)

%macro looper(val);

proc sql;

create table dx\_cnt.&val as

select

prov\_id,

sum(case

when substr(left(trim(&&dx.&val))1,1,3) ge ‘290’ and and substr(left(trim(&&dx.&val))1,1,3) le ‘319’ then 1

else 0

end as diag\_ind) as diag\_cnt.&val

from clm\_cnt

where begin\_dos between ‘01JAN2000’d a ‘31DEC2000’d

and substr(left(trim(&&dx.&val))1,1,3) ge ‘290’ and and substr(left(trim(&&dx.&val))1,1,3) le ‘319’

group by prov\_id

; quit;

%if &val =”1” %then %do;

data dx\_sum;

set dx\_cnt.&val;

run;

%else do;

proc append base = dx\_sum data=dx\_cnt.&val; run;

%end;

%mend looper;

%looper(val=1);

%looper(val=2);

%looper(val=3);

%looper(val=4);

%looper(val=5);

\*\*\* Final output data set =dx\_sum;

\*\*--- With arrays ---;

data dx\_cnt;

set work.claims (where = (begin\_dos between ‘01JAN2000’d a ‘31DEC2000’d));

array diag{5} dx1--dx5;

array diag\_cnt{5} dx\_cnt1--dx\_cnt5;

for i=1 to 5;

if substr(left(trim(diag1{i})),1,3) between 290 and 319 then diag\_cnt{i}=1;

else diag\_cnt{i}=0;

end;

\*\* Then summarize diag\_cnt1 - diag\_cnt5;

Proc sort data=dx\_cnt;

by prov\_id;

Proc summary data=dx\_cnt;

by prov\_id;

var diag\_cnt1 – diag\_cnt5;

output out =dx\_sum sum=;

run;

\*\*\* Final output = dx\_sum;

1. What SAS statement, within a data step, would be used to change the appearance of a date variable for a proc print?

format

1. Name some SAS procedures used to do data summarization.

proc summary

proc freq

proc sql

proc univariate

proc report

1. What SAS procedure is used to order SAS data sets? How can you remove duplicates?

proc sort, use nodupkey or nodup option

1. What SAS statement(s), can be used within a data step, to prevent the output of certain variables into the output data set?

keep/drop

1. What function do you use to grab a portion of a character variable?

substr

1. What SAS statement is used to allow a variable to not change its value from one data step iteration to the next?

retain

1. What function is used to convert a character variable to numeric?

input

1. What SAS system option(s) within a data step would be used to set the 9th and only the 9th observation from a data set?

if \_N\_ = 9

1. What SQL statement is necessary to build a SAS data set via proc sql?

proc sql;

**create table newData as**

select \*

from oldData

where (constraints)

; quit;

1. What SQL statement would be used to remove records from a SAS data set or dbms table?

delete

1. What SQL statement would be used to add records to a SAS data set or DBMS table?

union/union all

1. If I wanted to do a many to many join, can this be done with: a) only Proc SQL, b) only the merge statement or c) either a) or b)? Give an example of one to one match merging using the data step or proc sql.

Ans. c, both

proc sql;

create table myData as

select a.\*, b.\*

from data1 a

left join data2 b

on a.criteria = b.criteria

; quit;

proc sort data=data1; by criteria;

proc sort data=data2; by criteria;

data myData;

merge data1(in=a) data2(in=b);

by criteria;

if a or b;

run;

1. What SQL statement is used to change values in current records only?

case

1. What are the acceptable values for the first character of a macro variable definition? Please give an example.

letter (ex. %let m=1; ), underscore (%let \_m =1;)

1. What is the first character for all macro variables (being resolved)?

&

1. Name date formats you know in SAS and give examples.

* format date mmddyy10.; 01012001 for 01/01/2001
* format date date9.; 01JAN2001 for 01/01/2001
* format date mmddyy8.; 010101 for 01/01/01
* format date ddmmyy10.; 01012001 for 01/01/2001

1. What subroutine is used to assign a data set variable to a macro variable in the data step? Can I use this macro variable within the data step I created it?

call symput(‘var1’,var1);

1. If I were to resolve a macro variable within quotes, a) the quotes have to be single quotes, b) the quotes have to be double quotes, or c) either a) or b)?

B) Double

1. What macro statement writes text to the SAS log?

%put Some text here;