Using Macros to Compute US Health Insurance Coverage Estimates for Insertion into a Web-based Table Generator

Michele Burlew, *Episystems, Inc.*,
Peter Graven, Karen Turner, Michel Boudreaux, *University of Minnesota*

Objective

Develop a system of SAS® macro programs that manage the processes of:

1) estimating health insurance coverage in the US population based on responses to two federal surveys
2) formatting the results for loading into the SHADAC (State Health Access Data Assistance Center) Data Center.

This poster describes the steps followed to design, write, and test this specific system of SAS macro programs. These steps can be adapted when designing other systems.

Understand Data Sources

Health insurance estimates are derived from two federal surveys whose results are released annually in the fall:

- the US Current Population Survey's Annual Social and Economic Supplement (CPS), administered by the Bureau of Labor Statistics (BLS) and the Census Bureau
- the American Community Survey (ACS) administered by the Census Bureau.

Survey	Approximate Obs per Year	Data Center Estimates Start in:
CPS	200,000	1987
ACS	3,000,000	2008

Determine Variables Needed

The SAS SURVEYMEANS procedure computes the frequencies, percentages, and standard errors of the health insurance estimates for different descriptive characteristics and combinations of characteristics in the US population.

The health insurance measurements available in each survey were identified. Next, the variables that characterize the data were organized into two groups, the "Filter" variables and the "Row" variables.

Estimates are computed for subsets of the survey based on *filter variable* values. The values of *row variable* values are used to compute detail within subsets defined by the filter variables.

Organize the Processing Steps

Three major steps were identified in preparing the data for insertion into the SHADAC Data Center. Macro programs were written specifically for each step.

1. Prepare models and submit them to PROC SURVEYMEANS Produce one ODS output data set from each call to PROC SURVEYMEANS.
2. Combine all the output data sets from PROC SURVEYMEANS and post-process the data.
3. Create a CSV file from the output data set created in Step 2 to load into the web-based system.

Step 1: Analyses

- Determine computing environment
- Select variables
- Recode variables to create row and filter variables
- Run frequencies to validate recodes
- Run SURVEYMEANS models
- Save SURVEYMEANS output in data sets and PDF files
- Save SAS logs

Step 2: Combine

- Concatenate ODS output data sets from PROC SURVEYMEANS
- Post-process recoded variables and remove duplicate models
- Validate that all models are present

Step 3: CSVs

- Create single-year CSV from combined data set
- Calculate multi-year statistics from output
- Create multi-year CSV

Determine Analyses Needed

Models that contained valid combinations of the filter and row variables were determined. The code submits each model twice: once to analyze the US population and second to analyze each state using a BY STATE statement.

Model Category	Number of Unique Models Constructed	Total Models Analyzed Nationally and By State
Overall	1	52
Row Variables	17	884
Filter Variables x Row Variables	81	4,212
Filter Variables x Filter Variables x Row Variables	92	4,784
****Total**	**191**	**9,932**

This PROC SURVEYMEANS step analyzes 2009 CPS data by state. It analyzes the subpopulations defined by the two filter variables: AGEGRP3A and FPL2CATA and the one row variable AGEROW8A. Macro code generates the step by looping through lists of filter and row variables and their acronyms to construct the statements. Acronyms in the PDF filename and ODS output data set name identify the analysis.

```
ods pdf file="p:\data center\cps\results\2009\
     02 surveymeans\cps_sm_FFR_2009_20101111_abc1.pdf";
ods output domain=domainst_Fa3a_Fp2a_Rar8a;

proc surveymeans data=work.cps2009 stacking
                       mean nobs stderr sum;
  title "Processing CPS 2009";
  title2 "Filter AGEGRP3A X Filter FPL2CATA X Row
                AGEROW8A by State Model";
  by state;
  domain AGEGRP3A*FPL2CATA*AGEROW8A
         AGEGRP3A*FPL2CATA*AGEROW8A_I;
  cluster h_seq;
  strata geocode;
  var notcovered covered private empbased individual
      public medicaid medicare military;
  weight marsupwt;
  label AGEGRP3A= ' ' FPL2CATA= ' ' AGEROW8A= ' ' ;
run;
```

Develop Code in a Modular Style

Since this system of macros will analyze data annually following the release of the surveys, the code was written in a modular style to make it easier to handle changes in variable names and values that can occur over time and with changes in policy guidelines.

Macro program SEX_RECODE codes the row variable SEX in the CPS. (Variable SEX_I is needed to complete the model specification.)

```
%macro sex_recode;
  %if &cpsyear ge 1988 and &cpsyear ne 1995
                 %then %let sexvar=a_sex;
  %else %if &cpsyear=1995 %then %let sexvar=pesex;
  attrib sex label='Sex'
         sex_i label="SEX not missing";
  sex=&sexvar;
  if sex=. then sex=-1;
  sex_i=(sex not in (.,-1));
%mend sex_recode;
```

Modify Code in Steps of Increasing Complexity

Because the process of producing a file that can be loaded into the web-based system involves many steps, the macros were developed at first with minimal options. As experience was gained and output reviewed, more options were added.

```
%macro docpstables(startyear=,endyear=,cpsdataset=,
                   bystate=,dofreqs=yes) / minoperator;

%macro docpstables(startyear=,endyear=,cpsdataset=,
                   bystate=,dofreqs=yes,
                   dooverall=yes,dor=yes,dofr=yes,
                   doffr=yes,inits=) / minoperator;

%macro docpstables(startyear=,endyear=,cpsdataset=,
                   bystate=,dofreqs=yes,
                   doverall=yes,dor=yes,dofr=yes,
                   doffr=yes,
                   filterlist=all,rowlist=all,
                   inits=) / minoperator;
```

Write macros to execute with minimal intervention

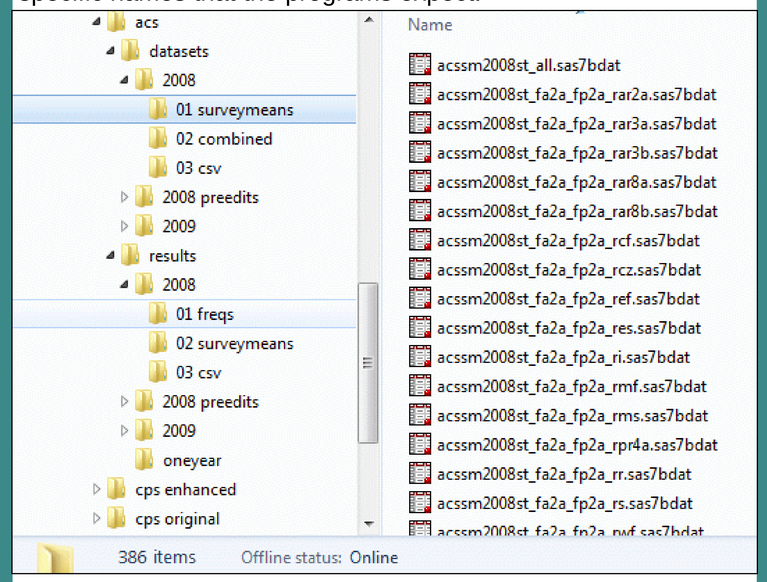
The SURVEYMEANS analyses take several days to complete. Manually submitting nearly 200 models per year and survey is time-consuming and error-prone. Macro code manages the model submissions. Loops iterated through the lists of filter variables and row variables to build the SURVEYMEANS steps and save the results.

These statements are from the section of the CPS processing program that constructs the Filter x Filter x Row models.

```
. . .
%do f=1 %to &nfiltervars;
  %do g=%eval(&f+1) %to &nfiltervars;
    %do r=1 %to &nrowvars;
      %let ffilter=%upcase(%scan(&filtervars,&f));
      %let gfilter=%upcase(%scan(&filtervars,&g));
      %let rowvar=%scan(&rowvars,&r);
. . .
      %let dsname=
F%scan(&filtercodes,&f)_F%scan(&filtercodes,&g)_R%scan(&ro
wcodes,&r);
      %let domainstmt=&ffilter*&gfilter*&rowvar
&ffilter*&gfilter*%scan(&rowmissvars,&r);
. . .
```

Devise Rules for Naming Files and Folders

Since the processing produces many output data sets and output files, the code uses pre-defined acronyms and survey identification to name files and output. Folders are defined with specific names that the programs expect.



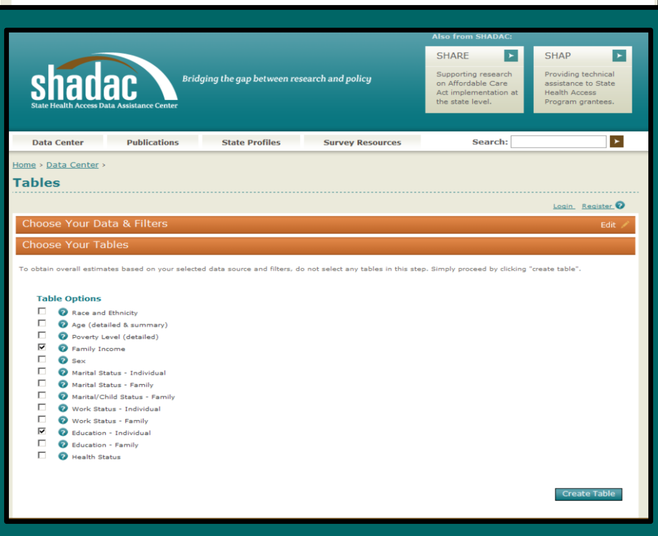
Save Logs and Output for Review

Because of the complexity and duration of the processing and the need to archive the results, the code saves SAS logs and procedure output in external files. PROC PRINTTO directs the SAS log to an external file. The ODS listing destination is closed and output is sent solely to a PDF file.

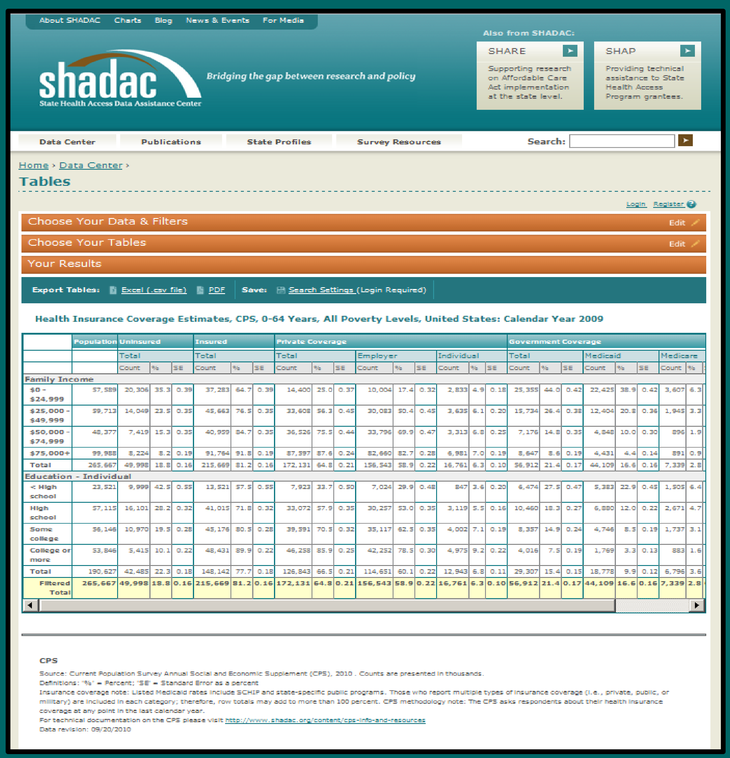
Screen Shots of SHADAC Data Center (www.shadac.org/datacenter)

The first screen that a user sees presents options to select the survey type, the geography, the year, and any age or poverty filter selections.

The second screen presents options to select the detail for the tables that contain the data specified in the first screen. One or more selections of the row variables can be made.



The third screen shows the tables of results. The output can be saved to a CSV or PDF file.



Acknowledgements

The SHADAC Data Center is funded by a grant from the Robert Wood Johnson Foundation.

Data processing to prepare the estimates for SHADAC's Data Center is conducted at the Minnesota Supercomputing Institute for Advanced Computational Research. (www.msi.umn.edu)

Contact Information

For SAS code questions, contact:
Michele Burlew
email: mmburlew@comcast.net

For SHADAC Data Center questions, contact:
SHADAC
2221 University Ave SE
Minneapolis, MN 55414
Phone: 612-624-4802 Fax: 612-624-1493
email: shadac@umn.edu