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, Minneapolis, MN

ABSTRACT

SAS® macro programs were developed to estimate health insurance coverage in the US population annually based on responses to two federal surveys and to format the results for loading into the SHADAC (State Health Access Data Assistance Center) Data Center. The SHADAC Data Center is a web-based table generator tool. (www.shadac.org/datacenter) The macro programs prepare data for analysis using the DATA step and base SAS procedures, analyze the data with PROC SURVEYMEANS, combine ODS output data sets from PROC SURVEYMEANS, and create CSV files that are loaded into the web-based system.

The SHADAC Data Center provides a way for health policy analysts to create customized tables and graphs of health insurance coverage estimates within a pre-defined set of parameters. The coverage estimates are derived from statistical analyses of the US Current Population Survey's Annual Social and Economic Supplement (CPS) and from the American Community Survey (ACS). These surveys are released annually in the fall.

The CPS analysis data set for one year contains approximately 200,000 observations. The ACS analysis data set for one year contains approximately 3 million observations. Both surveys are weighted to the population of the US. The SHADAC Data Center currently contains estimates for each year of CPS data from 1987 to 2009 and for ACS data from 2008 and 2009.

For each year and survey, macro programs code twenty-two classifications of the data such as age group, marital status, gender, and family income. The combinations of classifications generate almost 400 models for PROC SURVEYMEANS to analyze. Each model is processed nationally and by state.

On a PC with 4G of RAM using SAS 9.2, the 400 analyses for one year of the CPS data take approximately three hours, and the 400 analyses for one year of the ACS data take approximately 240 hours. Newly available to us is a Linux-based system that has SAS 9.2 where our testing shows that the ACS processing takes about one-third of the time that it does on a PC.

Several design considerations guided the development of the SHADAC Data Center SAS macros that can be applied when developing other similar applications. Writing code in a modular style and with the expectation that variable names and values will change over time make it easier to update code when changes in the data sources occur. When analyses generated by the macros take many hours to complete, defining macros that execute with minimal user intervention is necessary. Modifying code in steps of increasing complexity as output is verified and experience is gained ensures that the processing produces the required output. Saving all log and procedure output documents the processing for later review and archival purposes.

INTRODUCTION

The University of Minnesota's State Health Access Data Assistance Center (SHADAC) is funded by The Robert Wood Johnson Foundation to help states monitor rates of health insurance coverage, to understand factors associated with access to care, and to utilize data for implementation of health reform. SHADAC maintains an ongoing research agenda related to issues of health insurance coverage and access to care, data collection methods, and state health policy.

SHADAC began an effort in 2009 to build a data center where researchers and policy analysts could obtain tables and graphs of US and state health insurance coverage estimates from an online web-based table generator. The health insurance estimates are derived from two annual federal surveys. One survey is the US Current Population Survey's Annual Social and Economic Supplement (CPS). The CPS is administered as a joint effort of the Bureau of Labor Statistics (BLS) and the Census Bureau. The second survey is the American Community Survey (ACS), and the Census Bureau administers the ACS.

Both surveys randomly select respondents in the United States. The surveys obtain information about households, families, and persons. The survey questions cover many topics including demographics, income, education, and health insurance coverage. Analysts at BLS and the Census Bureau weight the surveys so that the responses reflect the US population as a whole.

The CPS has annual health insurance data from 1988 forward while the ACS has annual health insurance data from 2008 forward. Both surveys release their data in the fall. Each survey's federal web site has extensive tables and reports on various analyses of the data. The web sites also make the data and codebooks available by download so that analysts can conduct their own studies. The CPS data released fall 2010 was obtained from surveys made in March 2010. The ACS data released fall 2010 was obtained in 2009.

The goal of the initial release of the Data Center in fall 2009 was to provide health insurance coverage estimates for each year of the CPS (starting in 1988), and for each year of the ACS (starting in 2008). Additionally, SHADAC combined the CPS annual results to calculate two-year and three-year estimates. In fall 2011, the ACS plans to release a three-year file with responses combined from 2008, 2009, and 2010.

An important goal is to analyze the CPS and the ACS data as quickly as possible after the data and codebooks are released. Therefore, if SHADAC staff define new classifications since the previous year's release, these new classifications should be fully evaluated and the previous data releases completely re-analyzed before the fall release so that they can be incorporated in the upcoming fall release.

The SHADAC Data Center estimates up to nine different health insurance coverage types. Health insurance can be provided by public and private entities and some people have insurance from more than one source. The two surveys determine whether or not a person is covered by one or more sources. If a person is covered, the surveys determine whether the person's coverage is public, private, or a combination of the two. The surveys also determine the general types of public and private coverage that the respondent has.

The surveys employ different sampling strategies and measurement designs and therefore their health insurance estimates differ. By providing data from both surveys, SHADAC gives Data Center users the opportunity to compare results from both surveys and to work with data most appropriate to their needs.

Over 2,300 unique users have accessed the SHADAC Data Center since its start. The report "Cover Michigan. The State of Health Care Coverage in Michigan" produced by the Center for Healthcare Research and Transformation is an example of a document that was prepared using estimates from the Data Center.

The screenshots in Appendix 1 show the general layout of the online table generator.

DATA SOURCES

The records in each annual CPS data file are classified hierarchically into three different record types: household, family, and person. The CPS assigns household IDs, family IDs, and person IDs. These IDs link records across record types.

The ACS survey also has a hierarchical structure. It is released in two files; one file contains household information and the other file contains person-level information. As with the CPS survey, IDs link data across record types. SHADAC researchers add variables that define families within households.

Both surveys can have multiple families per household and multiple persons per family. A household can fit the common definition of a dwelling where one or more families reside. A household can also be just one individual who lives in an institution or group quarters like dormitories or barracks.

The 2009 CPS data file has about 400,000 records with data from about 97,000 households, 89,000 families, and 210,000 persons. Surveys from years prior to 2009 have similar numbers of records. With each CPS release, SHADAC researchers combine information from the different record types to create an analysis file that is at the person-level. Variables from the linked household and family records are added to the person records. The CPS analysis data set then has about 200,000 observations per year and it is at the person-level of the hierarchy.

The ACS survey in 2009 contained about 1 million records in the household file and 3 million records in the person-level file. As with each release of the CPS survey, SHADAC researchers combine the two ACS files so that the ACS analysis data set is at the person-level. Household and family variables are added to the person-level records. The ACS analysis data set has about 3 million observations.

STATISTICAL ANALYSES NEEDED

The SAS SURVEYMEANS procedure computes the frequencies, percentages, and standard errors of the health insurance estimates. Because of the features of the SAS programming language, it was decided to develop the processing in SAS using a combination of procedures, DATA steps, macro programs, and ODS output data sets.

In order to produce a consistent time-series, SHADAC staff produced a new set of weights for the CPS for each year and harmonized responses across questionnaire changes using imputation. Health insurance coverage estimates based on the original data as released by the federal government and based on the data with the SHADAC enhancements are available through the SHADAC Data Center. Further discussion of the enhanced weights is beyond the scope of this paper and information about them can be found on the SHADAC web site.

VARIABLE CONSTRUCTION

The goal of the SHADAC Data Center is to provide health insurance estimates for different descriptive characteristics and combinations of characteristics in the US population. The variables that characterize the surveys were organized into two groups. The first group of classification variables, the “filter variables”, provides Data Center users estimates based on subsets of the data. For example, a user may want to look at health insurance coverage in children only and not have the standard errors and percentages based on the population as a whole. The second group, the “row variables” provides detail within the population defined by the combination of filter variable values.

Some of the analyses were repeated. For example, the three age filter variables all have the same category of people 65 and older. Because of the requirements of the statistical analyses, it is not possible to omit the repeated analyses.

Table 1 lists the five filter variables as of the fall 2010 processing. The CPS and the ACS have the same set of filter variables. The acronyms were used to name the output data sets created by PROC SURVEYMEANS.

Table 1. SHADAC Data Center Filter Variables

Measurement	Variable Name	Acronym	Number of Values: Values
Age	AGEGRP2A	a2a	2: <65, 65+
Age	AGEGRP3A	a3a	3: 0-17, 18-64, 65+
Age	AGEGRP3B	a3b	3: 0-18, 19-64, 65+
Poverty Level Percentage	FPL2CATA	p2a	2: 0-<200, 200+
Poverty Level Percentage	FPL2CATB	p2b	2: 0-<139, 139+

Table 2 lists the row variables as of fall 2010. The HEALTHSTAT variable is unique to the CPS and it was first recorded in 1995. The CITIZENSHIP variable is unique to the ACS. The acronyms are used to name the output data sets created by PROC SURVEYMEANS.

Table 2. SHADAC Data Center Row Variables

Measurement	Variable Name	Acronym	Number of Values: Values
Age	AGEROW2A	ar2a	2: <65; 65+
Age	AGEROW3A	ar3a	3: 0-17; 18-64; 65+
Age	AGEROW3B	ar3b	3: 0-18; 19-64; 65+
Age	AGEROW8A	ar8a	8: 0-5; 6-17; 18-24; 25-34; 35-44; 45-54; 55-64, 65+
Age	AGEROW8B	ar8b	8: 0-5; 6-18; 19-24; 25-34; 35-44; 45-54; 55-64, 65+
Poverty level percentage	FPL4ROWCATA	pr4a	2: 0-<139; 139+
Race and ethnicity	RACE	r	5: Hispanic; White, non-Hispanic; Black, non-Hispanic; Asian; Multiple, non-Hispanic
Family income	FTOTCAT	i	4: $0-$24,999; $25,000-$49,999; $50,000-$74,999; $75,000+
Gender	SEX	s	2: Male; Female
Marital status of the individual who is 18 years of age or older	MARRIED	ms	2: Married; Not Married
Marital status of the family or household of the individual	FMAR	mf	2: Married; Not Married
Marital status and children under 18 in the family or household of the individual	MARWCHILD	cf	4: Married with children < 18; Married, without children < 18; Not married, with children < 18; Not married, without children < 18
Work status of the individual who is 18 years of age or older	WORKPRYRFT	ws	3: Not working; Part-time; Full-time
Work status of the family or household of the individual	FWORKPRYR	wf	3: None working; At least one part-time worker; At least one full-time worker
Education status of the individual who is 18 years of age or older	EDUCCAT	es	4: < High school; High-school; Some college; College or more
Highest education level in the family or household of the individual	FEDUCMAX	ef	4: < High school; High-school; Some college; College or more
Health Status (CPS only, starting in 1995)	HEALTHSTAT	h	2: Good/Very Good/Excellent; Poor/Very Poor
Citizenship (ACS only)	CITIZENSHIP	cz	2: Citizen; Non-citizen

The CPS also requires auxiliary data sources. One of these external sources is a poverty level lookup reference data set that identifies the poverty level for each year of the CPS based on ages and family size in the household. The other data sources needed by the CPS analyses are files of the revised weights that had been derived by SHADAC.

These revised weights provide an alternative to the weights that were supplied by the original release of the CPS. See the SHADAC Data Center web site for more information on these revised weights.

SAS PROGRAMMING DESIGN CONSIDERATIONS

Several design considerations guided the development of the SHADAC Data Center SAS code. The code was developed in a modular style because variable names and values can change over time. A modular style makes it easier to add and remove row and filter variables and to update the values of these variables when changes in the CPS or the ACS are made. Since the analyses take many hours to complete, and it is important to complete the analyses as quickly as possible, it was necessary to write the code so that it would execute with minimal intervention. Parameters to the macro programs were defined that kept these factors in mind. Because the Data Center would be widely available to researchers, saving all log and procedure output was necessary to document the processing.

Three major processing steps were identified for each year and survey type:

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 in a form that the vendor can load into the web-based system.

After determining these three major processing steps, code was developed separately for each step. Because of the large number of output files, folder structures reflected these three steps. Acronyms and filename standards were devised that identified the models, output data sets, and contents of logs and output files.

IDENTIFYING THE MODELS FOR SURVEYMEANS ANALYSES

After determining the set of filter variables and the set of row variables, the valid analyses that combine these variables were identified. Four different classes of models were defined and the number of models that needed to be constructed was determined. Each unique model is submitted twice: once to analyze the entire US population and once to analyze each state using a BY STATE statement. The BY STATE analyses analyze the fifty states and the District of Columbia.

Table 3 lists the four classes of models and the number of models constructed and analyzed. The model acronyms are used to name the logs and output files generated during the PROC SURVEYMEANS analyses. The table shows that there are 191 unique models. Each unique model is submitted twice: once to analyze the US as a whole and second to analyze each state using BY-variable processing. Considering the US model, the fifty states and DC, the total number of models analyzed per year and survey type is almost 10,000.

Some combinations of filter variables and row variables were omitted from the processing. For example, the filter variable AGEGRP3A crossed with the row variable AGEROW3B was not analyzed because of the mismatch between the first age category definitions of the two variables. The first level in AGEGRP3A selects persons age 0-17 while the first level in AGEROW3B selects persons age 0-18. A filter on the first level in AGEGRP3A would not include 18 year olds as defined by the first level in AGEROW3B. Analyzing a model that used these two filter variables would produce an incorrect analysis.

Table 3. SHADAC Data Center Model Categories and Types of Row Variables

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

With 191 models to construct and nearly 10,000 models to analyze on up to nine health insurance measures for each year and survey type, and because this processing would be repeated annually with each new survey release, it was necessary to devise a system of macros that would construct the models, submit the models, organize the output, and be reusable. Additionally, the processing time for some of the models can be several hours or days because of the size of the analysis data sets and complexity of the analyses. This time-consuming effort meant that it is necessary to operate this system of macros with minimal user intervention and to have the ability to review output at a later time to discover any problems in the processing. The code saves SAS logs and output in external files for later review and for archival purposes.

Before any processing begins, it is necessary to have a reference data set available that has an entry for all valid models and a second reference data set that has all possible combinations of the values produced by the valid models. These reference data sets only had to be created when there was a structural change in any of the filter or row variable definitions.

These reference data sets are used to ensure that all the required models are analyzed. Further, the vendor who manages the online data warehouse requires that the file produced for each year and survey type that they receive have one row for every possible combination of the values from the valid models. Some of the combinations might not actually occur in the survey data so this reference data set was used as a template on which to overlay the actual results. When a combination of the filter and row variables did not exist, the reference data set supplied a row of zeroes and missing values as appropriate. An example of a combination that would not actually occur in the data, but was required to be in the file sent to the vendor is when the age filter 0-18 is selected and the row value for age is 65+.

DATA CENTER PROCESSING STEP 1

The first step prepares the data for analysis, and constructs and submits the PROC SURVEYMEANS steps.

DATA CENTER PROCESSING STEP 2

The second step combines into one data set the nearly 400 output data sets per year and survey type that PROC SURVEYMEANS generates. (The fifty states plus DC are saved in one data set because of the use of the BY STATE statement. The US population results are saved in a second data set. Therefore, there are 191*2=382 output data sets per year and survey type.) The three filter variables for age are recoded into one variable as are the two filter variables for poverty. Similarly, the five row variables for age are recoded into one variable. After these recodes are made, PROC SORT with the NODUPREC option removes identical results. As mentioned in a previous section, for statistical reasons, it is necessary to fully specify a model even if part of it had already been analyzed.

Using the NODUPREC option instead of the NODUPKEY option was a way to ensure that there were no statistical result differences in models that analyzed different variables, but had some category definitions in common.

DATA CENTER PROCESSING STEP 3

The third step processes the data set created by the second step and creates a CSV file for the vendor to load into the Data Center data warehouse. Before outputting to the CSV, code sets statistic values to missing where the unweighted count for a combination is below a specific value. A DATA step merges this revised data set with the reference data set that contains all possible combinations of the data as described above.

DETERMINING THE SAS PROGRAM STRUCTURE

A similar set of macro programs was developed to analyze each of the two data sources. Although an overlap exists in how the filter variables and row variables are defined, the structures of the CPS and the ACS data sets is different enough that it was decided that it would be easier to maintain two different sets of programs, one for each of the two data sources.

FEATURES OF THE SAS PROGRAMS IN STEP 1

The program that completes Step 1 contains several macro program definitions. One of the macro programs initiates and manages the processing. This macro program defines several named parameters so that processing requests can be customized. The first actions in this driver macro program check the validity of the macro parameter values.

Several %LET statements at the beginning of the program define global macro variables whose values are paths for input and output. These macro variables are referenced throughout the processing.

The parameters to the driver macro program include:

- the range in years to process
- whether to analyze models for the entire US and/or for each state
- a list of specific filter variables to use in the analyses
- a list of specific row variables to use in the analyses
- the types of models to execute, such as “overall” or “row only”
- the platform on which to execute the models. The two choices are a PC or a Linux-based system.

The coding of the filter variables and row variables was done in a modular fashion with usually one macro program per variable. Variable names and variable values as assigned by the surveys tend to be the same over time. However, since these items might not remain consistent, the recoding macro programs started with statements that assigned data set variable names to macro variables. These macro variable names were then resolved in SAS language statements. The macro programs also might contain macro language statements to control processing based on the year of the survey.

An example of a recoding macro follows in Figure 1. Macro program SEX_RECODE codes the row variable SEX in the CPS. (Variable SEX_I is needed to complete the model specification in PROC SURVEYMEANS.)

Figure 1. Example of a Macro Program that Recodes a Row Variable

```
%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;
```

The calculations to compute the set of family variables is complex and requires several PROC and DATA steps to complete. Once the program completes coding the filter variables and row variables, table output can optionally be generated to verify the new variables.

The driver macro program then begins managing the submission of the models for analysis. One macro variable, FILTERVARS, stores the names of the variables that are specified as filter variables for the current execution of the driver macro program. Similarly, a second macro variable, ROWVARS, stores the names of the variables that are specified as row variables for the current execution. If the call to the driver macro program specifies only some of the filter and row variables, then code executes to manage the processing of these subsets of variables.

The model submission part of the driver macro program is organized by the four main categories listed in Table 3. The acronyms in Table 3 are used to name the logs and output files. There was one log file and one output file for each of the four categories of models. PROC PRINTTO directs the SAS log to an external file. Because of the volume of output, the listing destination is closed and output is sent solely to a PDF file.

The driver macro program constructs elements of the PROC SURVEYMEANS step by iterating through the names of the variables saved in macro variables FILTERVARS and ROWVARS. It skips over combinations that form invalid models. After defining several macro variables to insert into components of the PROC SURVEYMEANS step, it calls a macro program that finishes construction of the PROC SURVEYMEANS step and submits the step. Once the PROC SURVEYMEANS step and post-processing of the ODS output data set finish, control returns to the driver macro program to prepare the next submission to PROC SURVEYMEANS.

When the analyses for each model type finishes, PROC PRINTTO closes the log file and ODS closes the PDF output file associated with the model type.

Figure 2 shows an example of a PROC SURVEYMEANS step that analyzes 2009 CPS data by state. This step analyzes the subpopulations in each state that are defined by the two filter variables AGEGRP3A and FPL2CATA and the one row variable AGEROW8A. Acronyms in the name of the PDF file and in the name of the ODS output data set identify the analysis.

Appendix 2 shows excerpts of macro code. The first excerpt is from the driver macro program that constructs the elements of the PROC SURVEYMEANS step and calls the macro program that executes the PROC SURVEYMEANS step and processes its ODS output data set. The second excerpt is from the macro program that constructs the statements of the PROC SURVEYMEANS call.

Figure 2. Example of a PROC SURVEYMEANS Step

```
ods pdf nobackground 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;
```

FEATURES OF THE SHADAC DATA CENTER PROGRAMS IN STEP 2

When the processing for a survey type and a year finishes, the output data sets from PROC SURVEYMEANS are combined into one data set. This step executes quickly in a few minutes. A macro program obtains the names of the

SURVEYMEANS output data sets in the folder for the survey type and year by referencing the MEMBERS dictionary table in PROC SQL. The macro program uses %DO loops to concatenate the nearly 400 data sets into one data set for the survey type and year.

Once the full data set has been created, another macro program runs checks on the contents of the full data set. A PROC COMPARE step executes that compares the models present in the full data set to those in the template reference data set. Since the reference data set contains an observation for each model that should be analyzed, the PROC COMPARE step determines if output for a model is not present in the full data set. This data checking macro program also produces a report that lists counts that reflect specific checks of the full data set.

FEATURES OF THE SHADAC DATA CENTER PROGRAMS IN STEP 3

The last major step in the SHADAC Data Center processing creates a CSV file with PROC EXPORT for each year and survey type. The vendor who manages the Data Center data warehouse expects that the CSV file have a specific structure.

Before writing out the observations, this step rounds the percentages and standard errors. When counts are below a defined value, code sets the statistics to missing. This action prevents information that could potentially identify individuals from being displayed by the Data Center. The actual data is merged with the template reference data set so that an observation is present in the CSV file for all valid combinations of the filter and row variable values for each year and survey type.

PROCESSING TIME REQUIREMENTS

On a PC with 4G of RAM using SAS 9.2, the approximately 400 calls to PROC SURVEYMEANS for a full year's analysis took about 3 hours. As of the date of this paper, the processing for the ACS was split between the PC and a Linux-based system at the Minnesota Supercomputer Institute (MSI). The full analysis for the ACS took about 10 days of continuous processing. Newly available to SHADAC is a new Linux-based computer at MSI that has SAS 9.2 where our testing shows that the processing takes about one-third of the time.

CONCLUSION

Following the design guidelines of writing code in a modular style and developing it and testing it thoroughly in steps of increasing complexity can make it easier to write, maintain, and execute a complex system of programs. By following this path, SHADAC has been successful in achieving the goal of promptly estimating US health insurance coverage based on data from the CPS and the ACS soon after their annual survey data are released.

REFERENCES

ACS web site: http://www.census.gov/acs/www/

CPS web site: http://www.bls.gov/cps/

SHADAC Data Center web site: http://www.shadac.org/datacenter

SHADAC web site: http://www.shadac.org/

"Cover Michigan. The State of Health Care Coverage in Michigan". Center for Healthcare Research and Transformation. 2010. http://www.chrt.org/assets/cover-michigan/CHRT-Cover-Michigan-2010.pdf

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

Your comments and questions are valued and encouraged.

For SAS code questions, contact:

Michele Burlew
email: mmburlew@comcast.net

For SHADAC Data Center questions, contact:

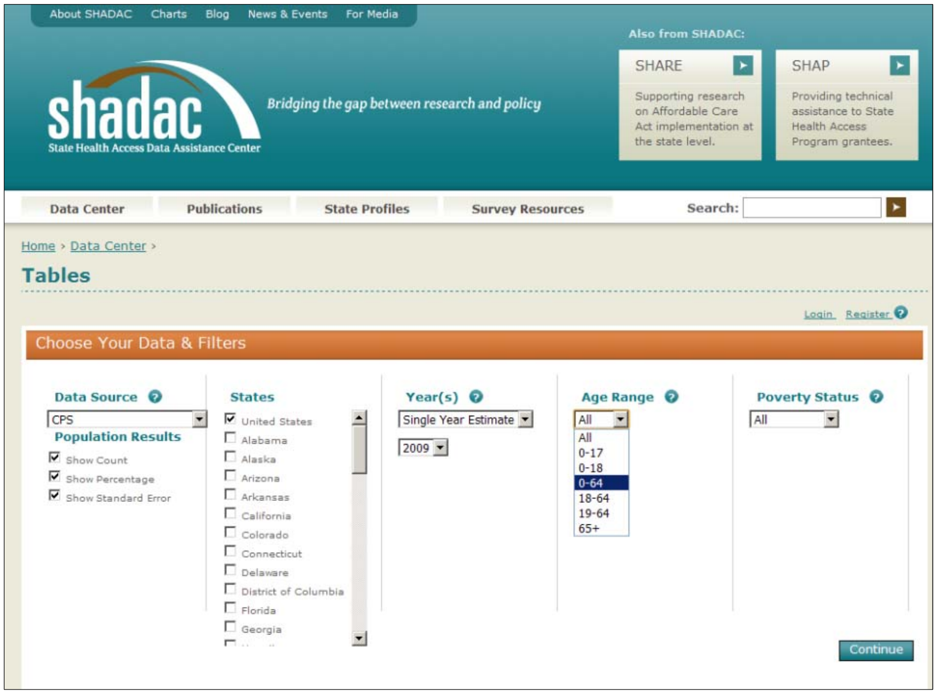
SHADAC
2221 University Ave SE
Minneapolis, MN 55414
Phone: 612-624-4802 Fax: 612-624-1493
email: shadac@umn.edu

APPENDIX 1

SCREEN SHOTS OF DATA CENTER

The following three screen shots show how SHADAC Data Center users can specify tables of health insurance estimates. These screen shots select tables of health insurance coverage estimates nationally in 2009 from the CPS for people less than 65 years of age and all poverty levels with detail by family income and by individual education level.

The first SHADAC Data Center screen presents options to select the survey type, the geography, the year, and any age or poverty filter selections.


About SHADAC
Charts
Blog
News & Events
For Media
Also from SHADAC:
SHARE
Supporting research on Affordable Care Act implementation at the state level.
SHAP
Providing technical assistance to State Health Access Program grantees.
shadac
State Health Access Data Assistance Center
Bridging the gap between research and policy
Data Center
Publications
State Profiles
Survey Resources
Search:
Home › Data Center ›
Login Register
Continue

Tables

Choose Your Data & Filters

Data Source: CPS

Population Results
- ☑ Show Count
- ☑ Show Percentage
- ☑ Show Standard Error

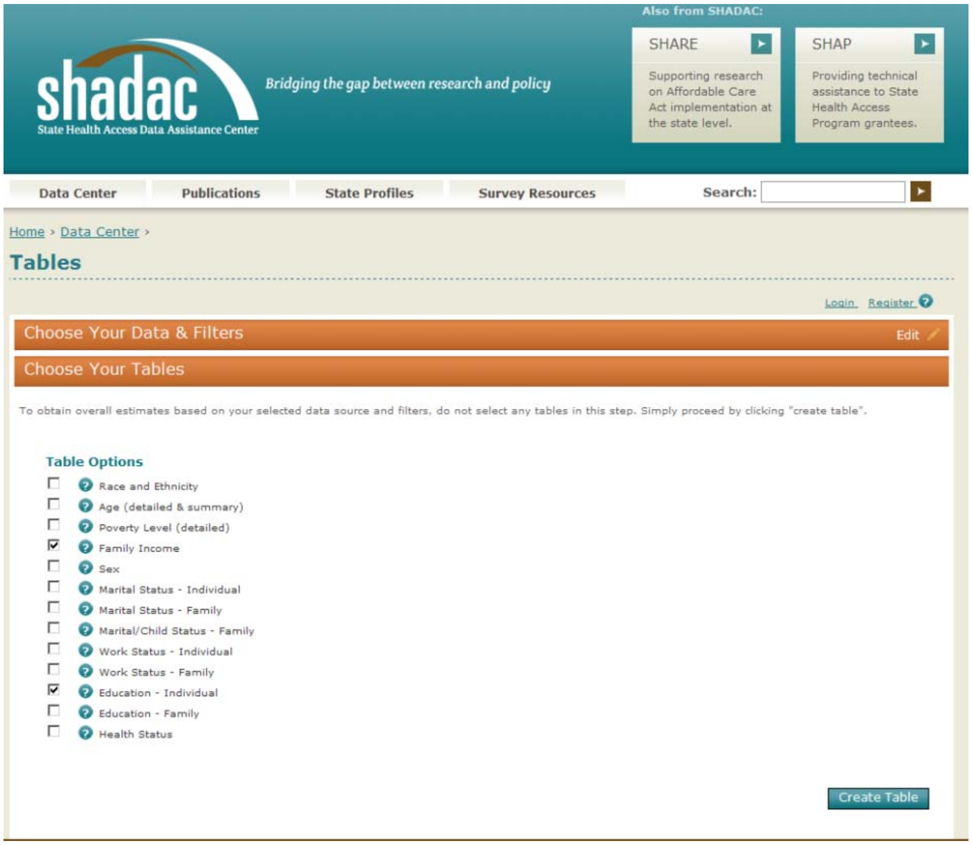
States
- ☑ United States
- □ Alabama
- □ Alaska
- □ Arizona
- □ Arkansas
- □ California
- □ Colorado
- □ Connecticut
- □ Delaware
- □ District of Columbia
- □ Florida
- □ Georgia

Year(s): Single Year Estimate; 2009

Age Range: All (All, 0-17, 0-18, 0-64 selected, 18-64, 19-64, 65+)

Poverty Status: All

The second SHADAC Data Center 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.


Also from SHADAC:
SHARE
Supporting research on Affordable Care Act implementation at the state level.
SHAP
Providing technical assistance to State Health Access Program grantees.
shadac
State Health Access Data Assistance Center
Bridging the gap between research and policy
Data Center
Publications
State Profiles
Survey Resources
Search:
Home › Data Center ›
Login Register
Edit
Create Table

Tables

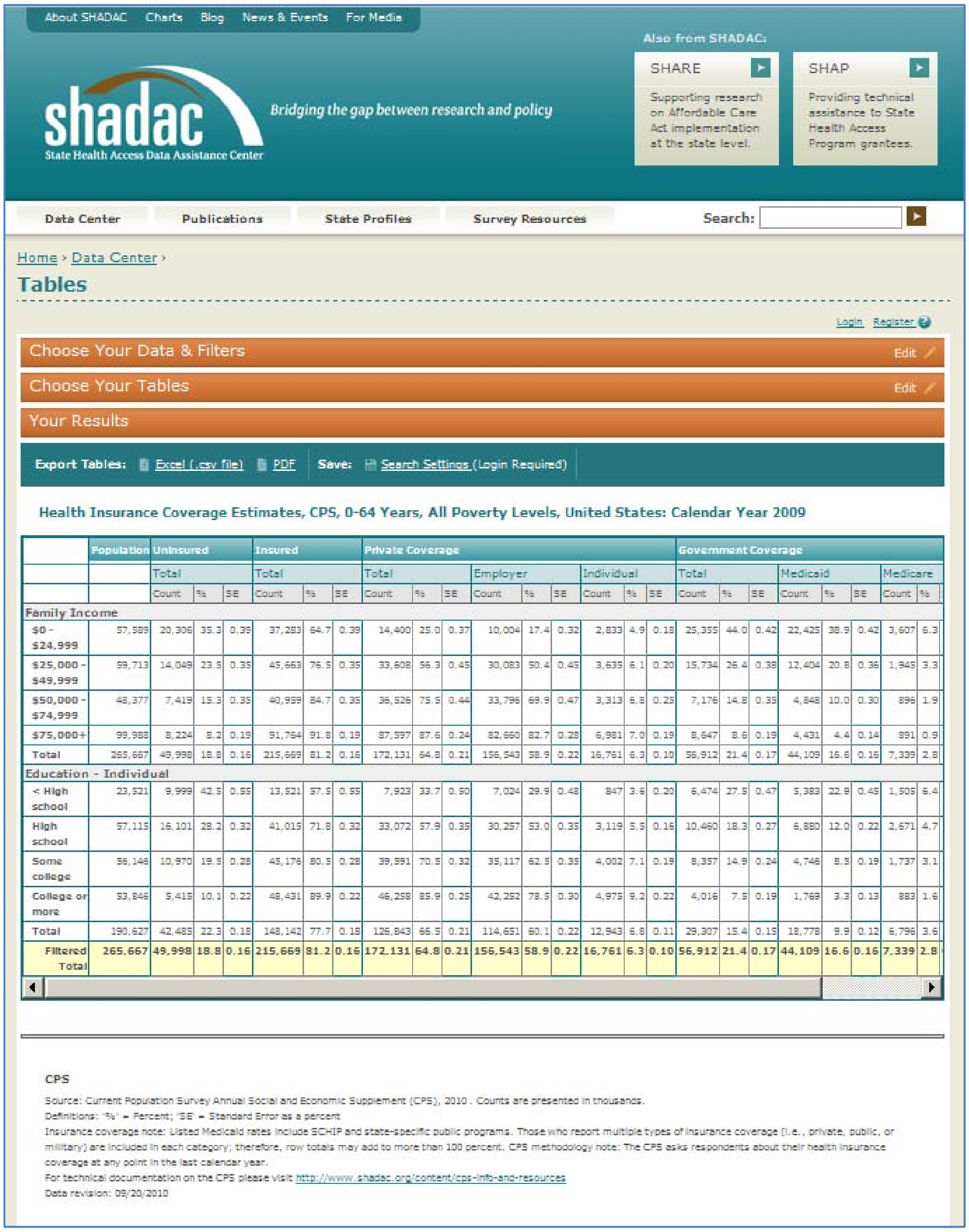
Choose Your Data & Filters

Choose Your Tables

To obtain overall estimates based on your selected data source and filters, do not select any tables in this step. Simply proceed by clicking "create table".

Table Options
- □ Race and Ethnicity
- □ Age (detailed & summary)
- □ Poverty Level (detailed)
- ☑ Family Income
- □ Sex
- □ Marital Status - Individual
- □ Marital Status - Family
- □ Marital/Child Status - Family
- □ Work Status - Individual
- □ Work Status - Family
- ☑ Education - Individual
- □ Education - Family
- □ Health Status

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



About SHADAC
Charts
Blog
News & Events
For Media
Also from SHADAC:
SHARE
Supporting research on Affordable Care Act implementation at the state level.
SHAP
Providing technical assistance to State Health Access Program grantees.
shadac
State Health Access Data Assistance Center
Bridging the gap between research and policy
Data Center
Publications
State Profiles
Survey Resources
Search:
Home › Data Center ›
Login
Register
Edit
Edit

Tables

Choose Your Data & Filters

Choose Your Tables

Your Results

Export Tables: Excel (.csv file) PDF Save: Search Settings (Login Required)

Health Insurance Coverage Estimates, CPS, 0-64 Years, All Poverty Levels, United States: Calendar Year 2009

	Population	Uninsured			Insured			Private Coverage									Government Coverage							
		Total			Total			Total			Employer			Individual			Total			Medicaid			Medicare	
		Count	%	SE	Count	%	SE	Count	%	SE	Count	%	SE	Count	%	SE	Count	%	SE	Count	%	SE	Count	%
Family Income																								
$0 - $24,999	57,589	20,306	35.3	0.39	37,283	64.7	0.39	14,400	25.0	0.37	10,004	17.4	0.32	2,833	4.9	0.18	25,355	44.0	0.42	22,425	38.9	0.42	3,607	6.3
$25,000 - $49,999	59,713	14,049	23.5	0.35	45,663	76.5	0.35	33,608	56.3	0.45	30,083	50.4	0.45	3,635	6.1	0.20	15,734	26.4	0.38	12,404	20.8	0.36	1,945	3.3
$50,000 - $74,999	48,377	7,419	15.3	0.35	40,959	84.7	0.35	36,526	75.5	0.44	33,796	69.9	0.47	3,313	6.8	0.25	7,176	14.8	0.35	4,848	10.0	0.30	896	1.9
$75,000+	99,988	8,224	8.2	0.19	91,764	91.8	0.19	87,597	87.6	0.24	82,660	82.7	0.28	6,981	7.0	0.19	8,647	8.6	0.19	4,431	4.4	0.14	891	0.9
Total	265,667	49,998	18.8	0.16	215,669	81.2	0.16	172,131	64.8	0.21	156,543	58.9	0.22	16,761	6.3	0.10	56,912	21.4	0.17	44,109	16.6	0.16	7,339	2.8
Education - Individual																								
< High school	23,521	9,999	42.5	0.55	13,521	57.5	0.55	7,923	33.7	0.50	7,024	29.9	0.48	847	3.6	0.20	6,474	27.5	0.47	5,383	22.9	0.45	1,505	6.4
High school	57,115	16,101	28.2	0.32	41,015	71.8	0.32	33,072	57.9	0.35	30,257	53.0	0.35	3,119	5.5	0.16	10,460	18.3	0.27	6,880	12.0	0.22	2,671	4.7
Some college	56,146	10,970	19.5	0.28	45,176	80.5	0.28	39,591	70.5	0.32	35,117	62.5	0.35	4,002	7.1	0.19	8,357	14.9	0.24	4,746	8.5	0.19	1,737	3.1
College or more	53,846	5,415	10.1	0.22	48,431	89.9	0.22	46,258	85.9	0.25	42,252	78.5	0.30	4,975	9.2	0.22	4,016	7.5	0.19	1,769	3.3	0.13	883	1.6
Total	190,627	42,485	22.3	0.18	148,142	77.7	0.18	126,843	66.5	0.21	114,651	60.1	0.22	12,943	6.8	0.11	29,307	15.4	0.15	18,778	9.9	0.12	6,796	3.6
Filtered Total	**265,667**	**49,998**	**18.8**	**0.16**	**215,669**	**81.2**	**0.16**	**172,131**	**64.8**	**0.21**	**156,543**	**58.9**	**0.22**	**16,761**	**6.3**	**0.10**	**56,912**	**21.4**	**0.17**	**44,109**	**16.6**	**0.16**	**7,339**	**2.8**

CPS

Source: Current Population Survey Annual Social and Economic Supplement (CPS), 2010 . Counts are presented in thousands.

Definitions: '%' = Percent; 'SE' = Standard Error as a percent

Insurance coverage note: Listed Medicaid rates include SCHIP and state-specific public programs. Those who report multiple types of insurance coverage (i.e., private, public, or military) are included in each category; therefore, row totals may add to more than 100 percent. CPS methodology note: The CPS asks respondents about their health insurance coverage at any point in the last calendar year.

For technical documentation on the CPS please visit http://www.shadac.org/content/cps-info-and-resources

Data revision: 09/20/2010

APPENDIX 2

CODE EXCERPT FROM DRIVER MACRO PROGRAM USED IN STEP 1

This section of code processes the filter variable by row variable models. It directs the log and output to external files. It cycles through the list of filter variables (FILTERVARS) and row variables (ROWVARS) that are to be analyzed and creates macro variables to hold the elements of the PROC SURVEYMEANS step. It skips over combinations of filter and row variables that form invalid models.

This section of code calls macro program DOCPSSTATS. The next code excerpt presents some of the code from DOCPSSTATS.

```
    %******************************************************;
    %* Do one filter variable by one row variable (doubles);
    %******************************************************;
  %double:
    %if &dofr=NO %then %goto triple;

    proc printto new
log="&dcpath\cps &cpsdataset\results\&cpsyear\02 surveymeans\
cps&cpsdataset.sm&modellevel._FR_&cpsyear._%sysfunc(today(),yymmddn8.)&inits..log";
    run;

    ods pdf
file="&dcpath\cps &cpsdataset\results\&cpsyear\02 surveymeans\
cps&cpsdataset.sm&modellevel._FR_&cpsyear._%sysfunc(today(),yymmddn8.)&inits..pdf"
style=minimal;

    %let starttime=%sysfunc(datetime());

    %let domainlist=;
    %let domainmisslist=;
    %let domainfilterlist=;
    %let domainstmt=;
    %let dsname=;
    %let smtitle=;

    %do f=1 %to &nfiltervars;
      %do r=1 %to &nrowvars;
        %let ffilter=%scan(&filtervars,&f);
        %let rowvar=%scan(&rowvars,&r);
        %if &cpsyear le 1995 and &rowvar=HEALTHSTAT %then %goto skiprowfr;

        %let domainlist=&ffilter &rowvar;

        %let domainmisslist=%scan(&filtermissvars,&f) %scan(&rowmissvars,&r);

        %if not(&ffilter=AGEGRP3A and &rowvar # AGEROW3B AGEROW2A) and
            not(&ffilter=AGEGRP3B and &rowvar # AGEROW3A AGEROW2A) %then %do;
          %let domainstmt=&ffilter*&rowvar &ffilter*%scan(&rowmissvars,&r);

          %let dsname=F%scan(&filtercodes,&f)_R%scan(&rowcodes,&r);
          %let smtitle=Filter &ffilter X Row &rowvar;

          %docpsstats
        %end;
        %else %put ******** Filter &ffilter X Row &rowvar skipped ********;

      %skiprowfr:
      %end;
    %end;

    %let endtime=%sysfunc(datetime());
    %let duration=%sysfunc(intck(minute,&starttime,&endtime));

    %put ** Started, Finished filter X row processing:
%left(%sysfunc(putn(&starttime,datetime25.)))
%left(%sysfunc(putn(&endtime,datetime25.)));
```

```
    %put ** Duration &duration minutes;

    proc printto;
    run;

    ods pdf close;

    %put ** Started, Finished filter X row processing:
%left(%sysfunc(putn(&starttime,datetime25.)))
%left(%sysfunc(putn(&endtime,datetime25.)));
    %put ** Duration &duration minutes;
```

CODE EXCERPT FROM MACRO PROGRAM THAT CONSTRUCTS PROC SURVEYMEANS STEP

Following is part of the code from macro program DOCPSSTATS that the driver macro program calls when processing the CPS. Before it calls DOCPSTATS, the driver macro program defines many of the elements that go into the PROC SURVEYMEANS step. These elements include items such as:

- the year of the analyses (macro variable CPSYEAR)
- name of output data set (macro variable DSNAME)
- whether to do national and state models (macro variable BYSTATE)
- the list of variables that go on the DOMAIN statement (macro variable DOMAINSTMT).

The %DO SMLOOP loop executes once or twice depending on whether national analyses only, state analyses only, or both types of analyses were specified on the invocation of the driver macro program.

```
%macro docpsstats;
  %if &cpsyear ge 1996 %then %do;
    %let hivars=
     notcovered covered private empbased individual public medicaid medicare military;
    %let nhivars=9;
  %end;
  %else %if &cpsyear ge 1988 and &cpsyear le 1995 %then %do;
    %let hivars=notcovered covered private public medicaid medicare military;
    %let nhivars=7;
   %end;

  %if &bystate=YES %then %do;
    %let startloop=1;
    %let smloops=2;
  %end;
  %else %if &bystate=NO %then %do;
    %let startloop=1;
    %let smloops=1;
  %end;
  %else %if &bystate=ONLY %then %do;
    %let startloop=2;
    %let smloops=2;
  %end;

  %* Second time through do it by states if BYSTATE=YES or ONLY;
  %do smloop=&startloop %to &smloops;
    %if &smloop=1 %then %let dsname=us_&dsname;
    %else %if &smloop=2 %then %do;
      %if %upcase(%substr(&dsname,1,2))=US %then %let dsname=st_%substr(&dsname,4);
      %else %let dsname=st_&dsname;
    %end;

    %if %quote(&domainstmt) ne %then %do;
      ods output domain=domain&dsname;
    %end;
    %else %do;
      ods output statistics=domain&dsname;
    %end;
    proc surveymeans data=cps&cpsyear stacking mean nobs stderr sum;
      title "Processing CPS &cpsdataset &cpsyear";
      %if &smloop=1 %then %do;
        title2 "&smtitle Model US";
      %end;
      %else %do;
        title2 "&smtitle by State Model";
        by state;
      %end;

      %if %quote(&domainstmt) ne %then %do;
        domain &domainstmt;
      %end;
      cluster h_seq;
      strata geocode;
      var &hivars;
      weight &cpsweight;

      %if %quote(&domainstmt) ne %then %do;
        label
        %let dv=1;
        %let domainvar=%scan(&domainlist,&dv);
        %do %until(&domainvar=);
          &domainvar=' '
          %let dv=%eval(&dv+1);
          %let domainvar=%scan(&domainlist,&dv);
        %end;
        ;
      %end;
    run;

    %let fullmember=cps%substr(&cpsdataset,1,1)sm&cpsyear&dsname;
    %let lfm=%length(&fullmember);
    %if &lfm gt 32 %then %let fullmember=%sysfunc(compress(&fullmember,_));

    %if &smloop=1 %then %do;
      data dc&cpsyear..&fullmember(label="CPS SURVEYMEANS &cpsyear Type &dsname US");
    %end;
    %else %do;
      data dc&cpsyear..&fullmember
               (label="CPS SURVEYMEANS &cpsyear Type &dsname by state");
    %end;
         . . . DATA step code omitted . . .

  %end;
%mend docpsstats;
```