

AGENCY FOR HEALTHCARE RESEARCH AND QUALITY

AHRQ

Analyzing MEPS-HC Data with SAS® 9.4M6

Pradip K. Muhuri

April 14-15, 2020

SAS® Steps

- SAS programs are primarily composed of one or more steps
  - DATA Step
  - PROC Step
- Common procedures used: PROC FREQ, PROC MEANS, PROC SUMMARY, PROC SORT, PROC DATASETS, PROC FORMAT, PROC PRINT, and PROC PRINTTO
- Complex survey procedures used: PROC SURVEYMEANS
- *SAS Windowing Environment* used as an interface to SAS to write, edit and submit SAS code
- Jupyter Notebook files for SAS exercises are also available here: https://github.com/pkmedu/Workshop and can be viewed from the browser directly

MEPS-HC SAS Transport Files on the Web

- https://meps.ahrq.gov/data_stats/download_data_files.jsp
- SAS Transport Files - best overall format
  - machine-independent **(**data files can be moved between computers running different operating systems).
  - can be directly imported into SAS, SPSS, BMDP, and STATA, etc.

Downloading MEPS-HC Data from the Web

- Objectives
  - Run SAS macro to automate the
    - downloading and unzipping of the SAS transport files
    - converting of the transport files into SAS data sets
  - Verify the following SAS data sets in your desired folder
    - 2017 MEPS Full-Year Consolidated File
    - 2016 MEPS Full-Year Consolidated File
    - 2017 Prescribed Medicine File
    - Panel 21 Longitudinal File
    - Panel 20 Longitudinal File
    - Panel 19 Longitudinal File

Exercise 1

- Objective
  - Generate the following estimates
    - mean health care expenses per person
    - mean health expenses per person with an expense (overall, and by age group)

- Data and Analysis
  - Use 2017 MEPS Full-Year Consolidated File
  - Run PROC FREQ for data checks
  - Run PROC SURVEYMEANS for complex survey estimates

MEPS Full-Year Consolidated File (Exercises 1, 2 & 3)

- This is a person-level data which includes annual variables such as
  - total annual healthcare expenditures by type of care
  - payment source, and type of provider seen
  - annual and monthly health insurance type indicators
  - health conditions, healthcare access and utilization
  - quality of care, patient satisfaction, and demographics

https://meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-201

- (For example) The 2017 Full-Year Consolidated File contains a total of 31,880 persons who were part of one of the two MEPS panels for whom data were collected in that year:
  - 2017 portion of Round 3, Rounds 4 and 5 for Panel 21
  - Rounds 1, 2 and the 2017 portion of Round 3 for Panel 22

PROC SURVEYMEANS

- The MEPS-HC uses sample design features including stratification, clustering, and oversampling.
- Due to complexities in the MEPS-HC sample designs (https://meps.ahrq.gov/data_files/publications/mr33/mr33.shtml), we must specify the survey weight and design characteristics in PROC SURVEYMEANS step when estimating the parameter for the U.S. civilian noninstitutionalized population.
- The following is an example SAS program.

(SAS/STAT® 15.1 User's Guide Introduction to Survey Sampling and Analysis Procedures https://support.sas.com/documentation/onlinedoc/stat/151/introsamp.pdf)

```
PROC SURVEYMEANS DATA=WORK.PUF201;
        VAR  TOTEXP;
        STRATUM VARSTR;
        CLUSTER VARPSU;
        WEIGHT PERWT17F;
        RUN;
```

Notes: If you do not specify statistic-keywords in the PROC SURVEYMEANS statement, it computes the NOBS, MEANS, STDERR, and CLM statistics by default. When you request SUM (i.e., estimated population total when the appropriate sampling weights are used), the procedure computes STD by default.

PROC SURVEMEANS (Statements Explained)

```
PROC SURVEYMEANS DATA=WORK.PUF201;
        VAR  TOTEXP ;
        STRATUM VARSTR;
        CLUSTER VARPSU;
        WEIGHT PERWT17F;
RUN;
```

VAR identifies the variable to be analyzed.

STRATUM lists the variable that form the strata.

CLUSTER specifies the cluster identification variable.

WEIGHT names the sampling weight variable.

PROC SURVEYMEANS (DOMAIN Statement Explained– Exercise 1)

```
ODS EXCLUDE STATISTICS; /* Not to generate output for the overall
population */
PROC SURVEYMEANS DATA= WORK.PUF201 ;
        VAR  TOTEXP;
        STRATUM VARSTR ;
        CLUSTER VARPSU ;
        WEIGHT  PERWT17F ;
        DOMAIN X_ANYSVCE('1')  X_ANYSVCE('1')*AGECAT ;
        FORMAT  AGECAT agecat.;
RUN;
```

Example on Domain Specs: The specification X_ANYSVCE('1')*AGECAT indicates that only the results associated with X_ANYSVCE=1 (subpopulation) for each category of AGECAT are of interest here.

Exercise 2

- Objective
  - Estimate the following parameters
    - mean number of purchases of opioids (i.e., Narcotic analgesics or Narcotic analgesic combos) per person with one or more purchases of opioids
    - mean total, out-of-pocket, and third-party payer expenses for purchases of opioids per person with one or more purchases of opioids
- Data and Analysis
  - Aggregate 2017 MEPS prescribed medicines data at the person-level
  - Merge aggregated prescribed medicine data with full-year person-level data for the same year
  - Run PROC FREQ for data checks
  - Run PROC SURVEYMEANS for complex survey estimates

MEPS Prescribed Medicines File (Exercise 2)

- This file contains prescribed medicine records.
- Each record represents one household-reported prescribed medicine that was purchased during a given calendar year.
- The variables include DRUGIDX, DUPERSID, LINKIDX, expenditures and source of payment, etc. Click this link for the list of all variables.

https://meps.ahrq.gov/data_stats/download_data/pufs/h197a/h197adoc.pdf

- The data included in the 2017 file were collected during the
  - 2017 portion of Round 3, Rounds 4 and 5 for Panel 21
  - Rounds 1, 2 and the 2017 portion of Round 3 for Panel 22

Exercise 3

- Objective
  - Estimate mean out-of-pocket health care expenses for individuals who were aged 26-30 years with high income and uninsured for the whole year

- Data and Analysis
  - Combine data from 2016 and 2017 MEPS Full-Year Consolidated Files
  - Run PROC FREQ and PROC MEANS for data checks
  - Run PROC SURVEYMEANS for complex survey estimates

Exercise 4

- Objective
  - Estimate the percentage distribution of insurance status (in the second year) of individuals who were aged 26-30 with high income and uninsured for the whole (first) year

- Data and Analysis
  - Combine data from MEPS Longitudinal Files (Panels 19, 20, and 21)
  - Run PROC FREQ and PROC MEANS for data checks
  - Run PROC SURVEYMEANS for complex survey estimates

MEPS Longitudinal File (Panel 21 as an Example) - Exercise 4

- This file is a two-year longitudinal file derived from the respondents to the MEPS Panel 21 sample. The persons on this data set represent those who were in the MEPS population (U.S. civilian noninstitutionalized) for all or part of the 2016-2017 period.

- The file contains a longitudinal weight variable (LONGWT) and all variables from the 2016 and 2017 full-year consolidated data files (HC-192 and HC-201, respectively).

https://meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-202

- The weight variable (LONGWT), when applied to the persons who participated in both 2016 and 2017, will enable the user to make national estimates of person-level changes in selected variables (e.g., health insurance, health status, utilization and expenditures).

PROC SURVEYMEANS
(CLASS Statement Explained – Exercise 4)

```
ODS EXCLUDE STATISTICS; /* Not to generate output for the overall population */
/* PROC SURVEYMEANS computes the NOBS, MEANS, STDERR, and CLM statistics
by default */
PROC SURVEYMEANS DATA=POOL;
        VAR  INSCOVY2;
         STRATUM VARSTR ;
         CLUSTER VARPSU ;
         WEIGHT  POOLWT;
         CLASS INSCOVY2;
         DOMAIN  SUBPOP("AGE 26-30, UNINSURED WHOLE YEAR, AND HIGH
INCOME");
         FORMAT INSCOVY2 INSF. SUBPOP SUBPOP.;
RUN;
```

Note: With PROC SURVEYMEANS, if you want categorical analysis for a numeric variable, you must include that variable in both VAR and CLASS statements.