U.S. Department of Health and Human Services
Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

MEPS Data Tools and Programming Overview

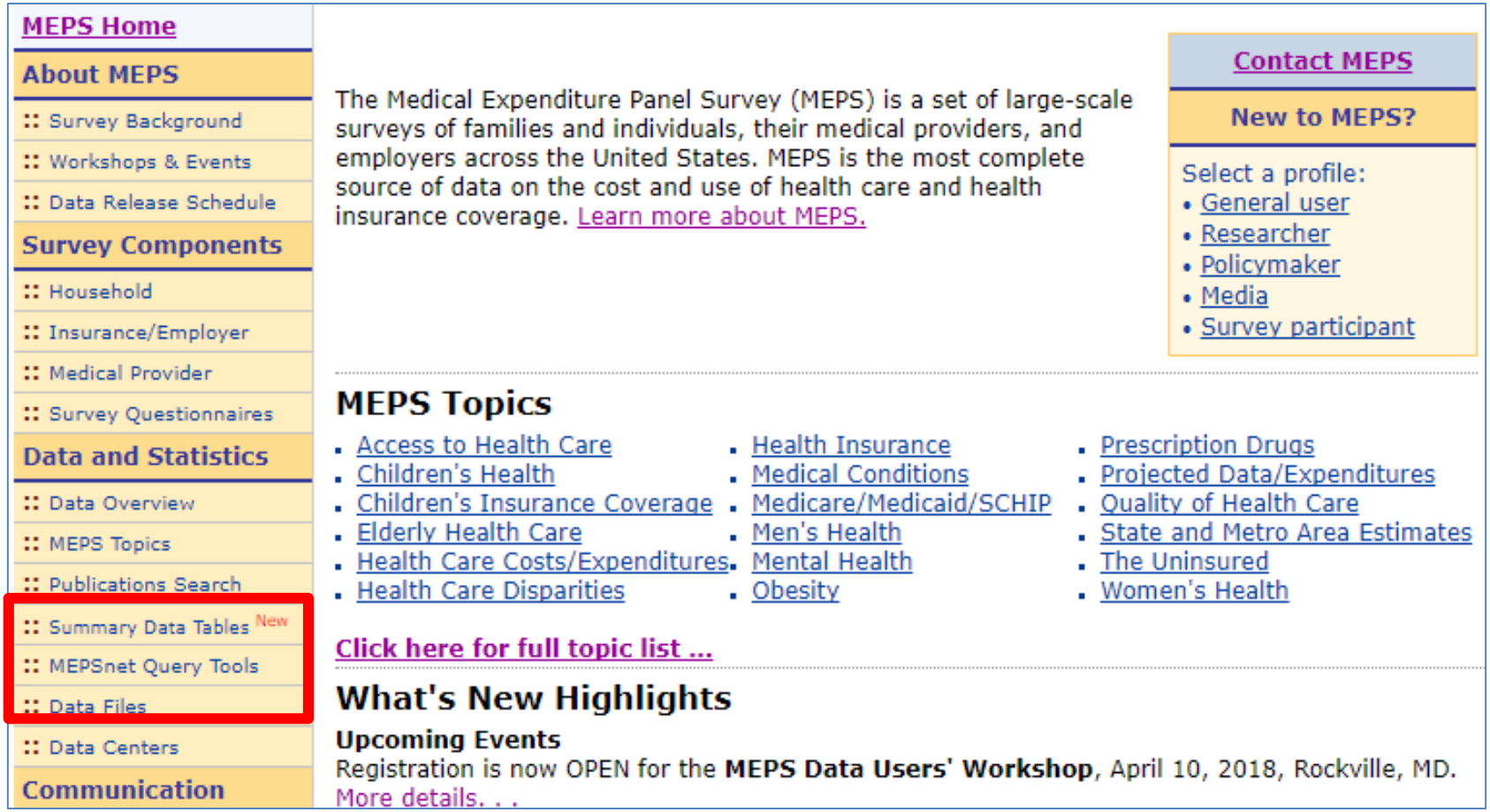
Emily Mitchell, PhD

Table of contents

Data Tools

MEPS Data Tools

www.meps.ahrq.gov



Summary Data Tables

new

MEPS summary tables

Household Component Tables | MEPS Home

Household Component summary tables

The MEPS Household Component summary tables provide frequently used summary estimates for the U.S. civilian noninstitutionalized population on household medical utilization and expenditures, demographic and socio-economic characteristics, health insurance coverage, access to care and satisfaction with care, medical conditions, and prescribed medicine purchases. Most tables can be stratified by demographic or socio-economic characteristics. Plots from selected data can also be generated, and R and SAS code for calculating selected estimates is available. See Sample Design and Data Collection Process for details on the collection of individual data items (e.g., health insurance status, age). The estimates provided in the tables are based on data available in standardized public use data files. Pages have been optimized for Chrome, Firefox, and Safari.

Use, expenditures, and population

Utilization, spending, and population totals by demographic characteristics, event type, or source of payment.

Health insurance

Number and percentage of people by insurance coverage and demographic characteristics.

Accessibility and quality of care

Number and percentage of people with a usual source of care, difficulty accessing needed care, preventive care, diabetes care, and patient-reported quality of doctor's visits, by demographic characteristics.

Summary Data Tables

new

Health insurance

These MEPS summary tables provide statistics on health insurance coverage for all ages, persons under 65, and those 65 and older. Data can be viewed over time or for a single year by demographic characteristics (such as age, race, or sex).

Use the options below to select a statistic (number or percentage of people), variable of interest (insurance coverage category), data view ("Trends over time" or "Cross-sectional"), and data years. If you select "Trends over time", you can choose a range of years. The "Cross-sectional" view displays a single year, which can be stratified by a grouping variable. Once a grouping variable is selected, a dropdown will appear, enabling selection of specific levels in each group.

After you select the available options, the table will automatically be updated. The data can be viewed as a plot under the "Plot" tab, with line graphs for trends over time and grouped bar graphs for the cross-sectional view. The "Code" tab displays R and SAS code needed to replicate the data shown in the table. The generated table, plot, and codes can be downloaded with the download button under each tab. To view standard errors in the table or 95% confidence intervals in the plot, select the "Show standard errors" checkbox.

Select statistic:

Number of people

Show standard errors

Select variable:

Insurance coverage, all ages

Select Levels

Select data view:

Trends over time

Cross-sectional

Year: **to:**

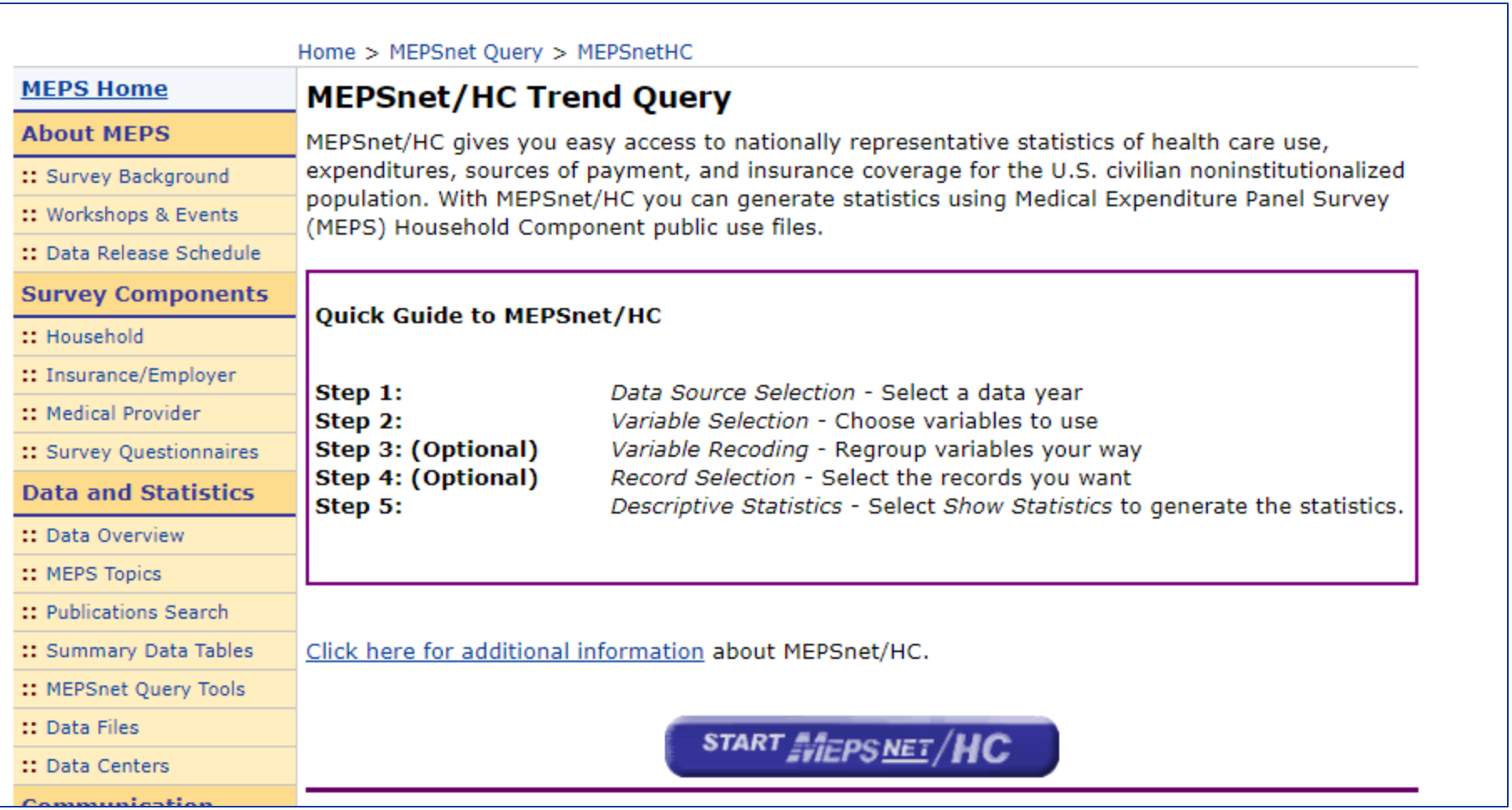
1996 2015

Table | Plot | </> Code

Number of people in thousands by insurance coverage, all ages, United States, 1996-2015

Year	Any private, all ages	Public only, all ages	Uninsured, all ages
2015	214,446	80,828	26,149
2014	208,377	78,739	31,324
2013	201,609	73,576	40,537
2012	201,911	71,733	39,847
2011	203,056	69,113	38,957
2010	200,580	67,557	40,437
2009	201,395	63,769	41,497
2008	201,723	61,824	40,828
2007	201,886	59,272	40,151

MEPSnet Query Tools



Public Use Files

MEPS Public Use Data Files

MEPS Public Use Data Files

- :: Data Centers
- **Communication**
- :: What's New
- :: Mailing List
- :: Discussion Forum
- :: Participants' Corner

Update notes

Documentation	File type
Documentation	PDF (500 KB) / HTML
Codebook	PDF (121 KB) / HTML*
SAS Programming Statements	TXT (59 KB)
SPSS Programming Statements	TXT (343 KB)
STATA Programming Statements	TXT (348 KB)

Data	File type**
Data File, ASCII format	ZIP (8.3 MB) / EXE (8.8 MB)
Data File, SAS transport format	ZIP (9.3 MB) / EXE (9.9 MB)

Questionnaires — see **Survey Questionnaires**

*The PDF version of the codebook is recommended for printing; the HTML version is database driven and lets you navigate quickly to details on each variable.
**Right-click on the data file link, then select Save Target As or Save Link As to download the file.

MEPS Public Use Data Files

- :: Data Centers
- **Communication**
- :: What's New
- :: Mailing List
- :: Discussion Forum
- :: Participants' Corner

Update notes

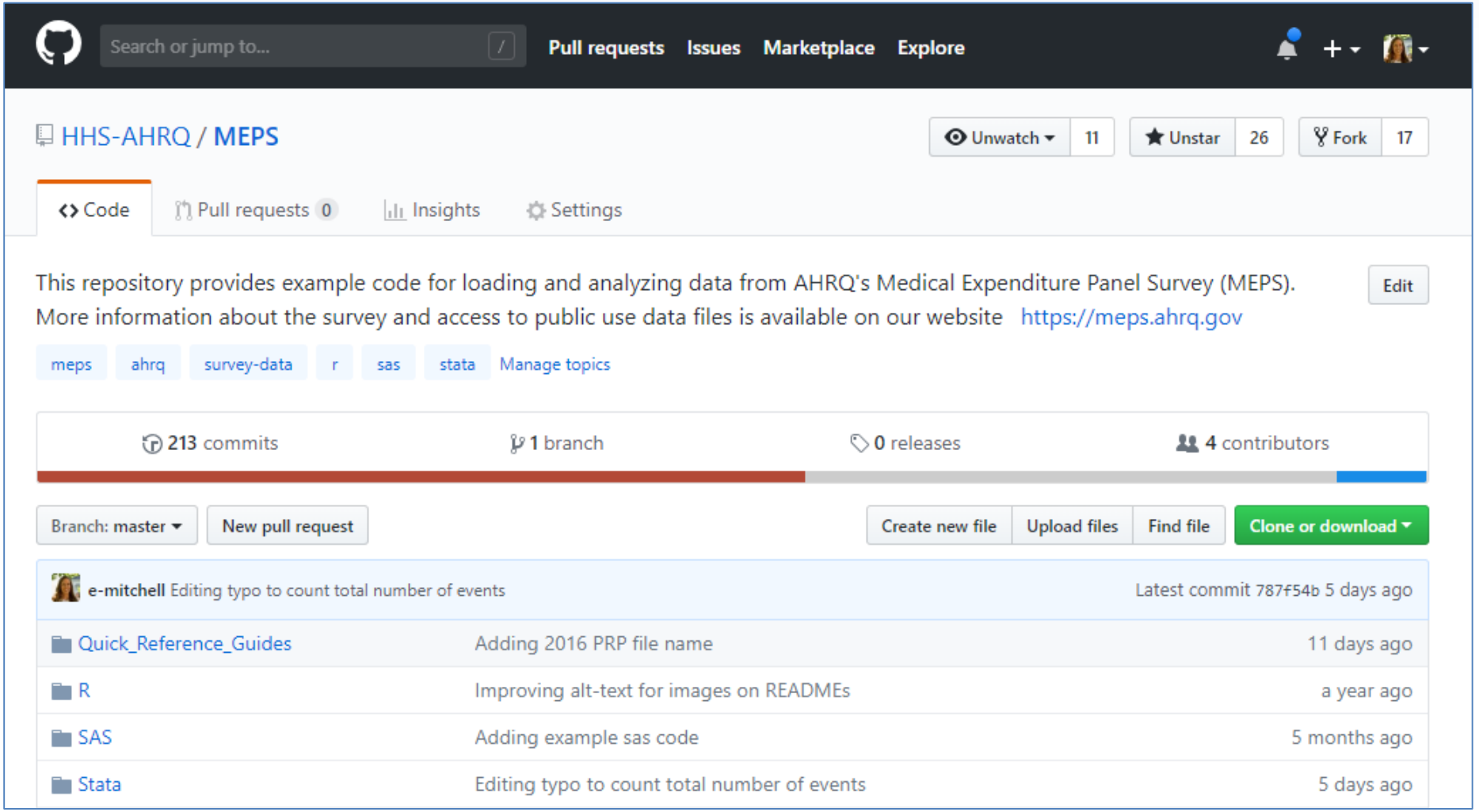
Documentation	File type
Documentation	PDF (500 KB) / HTML
Codebook	PDF (121 KB) / HTML*
SAS Programming Statements	TXT (59 KB)
SPSS Programming Statements	TXT (343 KB)
STATA Programming Statements	TXT (348 KB)

Data	File type**
Data File, ASCII format	ZIP (8.3 MB) / EXE (8.8 MB)
Data File, SAS transport format	ZIP (9.3 MB) / EXE (9.9 MB)

Questionnaires — see **Survey Questionnaires**

*The PDF version of the codebook is recommended for printing; the HTML version is database driven and lets you navigate quickly to details on each variable.
**Right-click on the data file link, then select Save Target As or Save Link As to download the file.

HHS-AHRQ GitHub



https://github.com/HHS-AHRQ/MEPS

Record-Level and Identifiers

Person-level

- Full-year consolidated file
- Longitudinal files

Event-level

- Event files: RX, DN, OM, IP, ER, OP, OB, HH

Condition-level

- Medical conditions file

Job-level

- Jobs file

Record-Level and Identifiers

Person-level

DUID	PID	DUPERSID
20004	101	20004101
20004	102	20004102
20004	103	20004103
20005	101	20005101

Event-level

DUPERSID	EVNTIDX
20004101	200041010011
20004101	200041010021
20005101	200051010151
20005101	200051010201

Conditions-level

DUPERSID	CONDN	CONDIDX
20004103	11	200041030011
20005101	11	200051010011
20005101	21	200051010021
20005101	51	200051010051

Jobs-level

DUPERSID	RN	JOBSN	JOBSIDX
20004101	3	1	2000410130 1
20004101	4	1	20004101401
20004101	5	1	20004101501
20004102	3	1	20004102301

Record-Level and Identifiers

Person-level

DUID	PID	DUPERSID
20004	101	20004101
20004	102	20004102
20004	103	20004103
20005	101	20005101

Event-level

DUPERSID	EVNTIDX
20004101	200041010011
20004101	200041010021
20005101	200051010151
20005101	200051010201

Conditions-level

DUPERSID	CONDN	CONDIDX
20004103	11	200041030011
20005101	11	200051010011
20005101	21	200051010021
20005101	51	200051010051

Jobs-level

DUPERSID	RN	JOBSN	JOBSIDX
20004101	3	1	20004101301
20004101	4	1	20004101401
20004101	5	1	20004101501
20004102	3	1	20004102301

Record-Level and Identifiers

Person-level

DUID	PID	DUPERSID
20004	101	20004101
20004	102	20004102
20004	103	20004103
20005	101	20005101

Event-level

DUPERSID	EVNTIDX
20004101	200041010011
20004101	200041010021
20005101	200051010151
20005101	200051010201

Conditions-level

DUPERSID	CONDN	CONDIDX
20004103	11	200041030011
20005101	11	200051010011
20005101	21	200051010021
20005101	51	200051010051

Jobs-level

DUPERSID	RN	JOBSN	JOBSIDX
20004101	3	1	20004101301
20004101	4	1	20004101401
20004101	5	1	20004101501
20004102	3	1	20004102301

Variable Naming Conventions

Edited Variables end in an “X”

RACE**X**

Year-specific variables use last two digits of year

TOTEXP**16**
PERWT**16**F

Round-specific variables, use two-digit round

- Some questions only asked in certain rounds, e.g. the Self-Administered Questionnaire in rounds 2 and 4

AGE**31**X
AGE**42**X
AGE**53**X

Estimation Variables

Weight Variables

- Person-level (e.g. PERWT16F, DIABW16F, SAQWT16F)
- Family-level (e.g. FAMWT16F, FAMWT16C)
- Longitudinal (e.g. LONGWT)

Variance-Estimation Variables:

- After 2002 FY data: VARSTR, VARPSU
- 1996-2001 FY data: VARSTRyy, VARPSUyy
  - When calculating variances with ***pooled data***, use STRA9616, PSU9616 in HC-036

MEPS Reserve Codes

-1	Inapplicable	Question was not asked due to skip pattern
-7	Refused	Question was asked and respondent refused to answer question
-8	Don’t Know	Question was asked and respondent did not know answer
-9	Not Ascertained	Interviewer did not record the data
-10	Top-Coded	Variable was top-coded for confidentiality

MEPS Reserve Codes

-1 Inapplicable

-7 Refused

-8 Don't Know

-9 Not Ascertained

-10 Top-Coded ⟵ Jobs file

EXAMPLES

FYC file: Pregnancy

Event file: Expenditures for phone calls

Jobs file: Hourly Wage

Programming Overview

Programming Example

How do 2016 medical expenses for the elderly (age 65 and over) compare to those for persons under 65?*

* Not including people that have $0 in expenses

Process

1. Load datasets
2. Create new variables (if needed)
3. Run survey procedure
4. Examine results

Process

1. **Load datasets**
2. Create new variables (if needed)
3. Run survey procedure
4. Examine results

Load datasets

- :: Data Centers
- **Communication**
- :: What's New
- :: Mailing List
- :: Discussion Forum
- :: Participants' Corner

Update notes

Documentation	File type
Documentation	PDF (500 KB) / HTML
Codebook	PDF (121 KB) / HTML*
SAS Programming Statements	TXT (59 KB)
SPSS Programming Statements	TXT (343 KB)
STATA Programming Statements	TXT (348 KB)

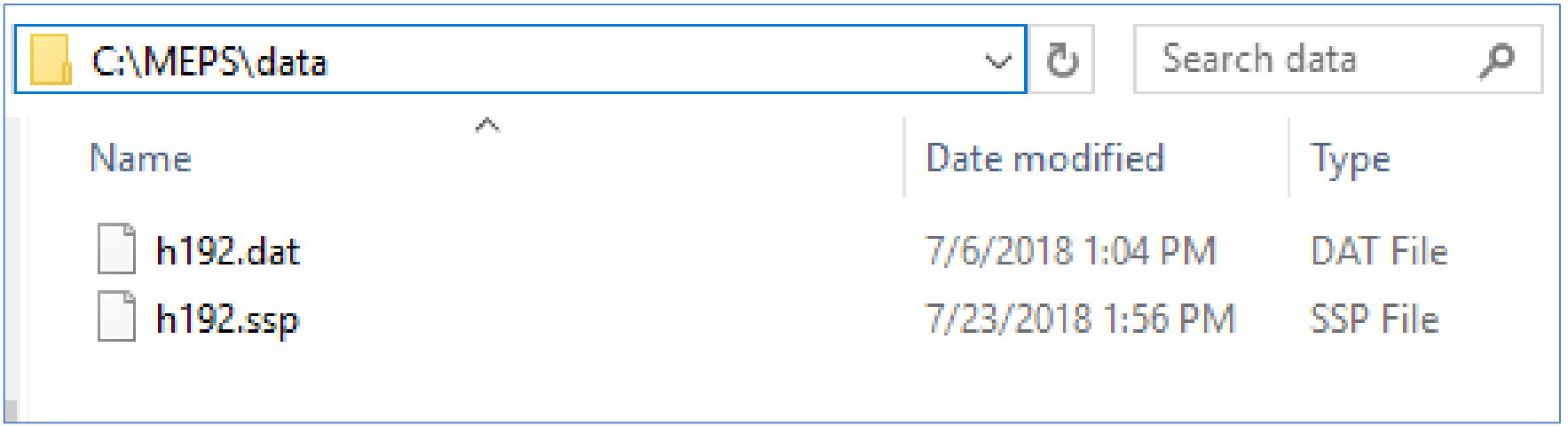
Data	File type**
Data File, ASCII format	ZIP (8.3 MB) / EXE (8.8 MB)
Data File, SAS transport format	ZIP (9.3 MB) / EXE (9.9 MB)

Questionnaires — see **Survey Questionnaires**

*The PDF version of the codebook is recommended for printing; the HTML version is database driven and lets you navigate quickly to details on each variable.
**Right-click on the data file link, then select Save Target As or Save Link As to download the file.

Load datasets

Store .dat and .ssp files in a local directory:



Load datasets

SAS

```
FILENAME in_h192 'C:\MEPS\data\h192.ssp';

proc xcopy in = in_h192 out = WORK IMPORT;
run;
```

Stata

```
import sasxport "C:\MEPS\data\h192.ssp"
```

R

```
install.packages("foreign"); library(foreign);

h192 = read.xport("C:/MEPS/data/h192.ssp")
```

Load datasets -- shortcut

SAS

```
%load_MEPS(h192);
```

Stata

```
copy "https://meps.ahrq.gov/mepsweb/data_files/pufs/h192ssp.zip" "h192ssp.zip"
unzipfile "h192ssp.zip"

import sasxport "h192.ssp", clear
```

R

```
install.packages("foreign"); library(foreign);

url = "https://meps.ahrq.gov/mepsweb/data_files/pufs/h192ssp.zip"
download.file(url, temp <- tempfile())

h192 = read.xport(unzip(temp))
unlink(temp)  # Unlink to delete temporary file
```

Process

1. Load datasets
2. **Create new variables (if needed)**
3. Run survey procedure
4. Examine results

Create new variables

SAS

```
data h192;
  set h192;

  if 0 <= AGELAST <= 64 then agecat = 1;
  else if AGELAST > 64 then agecat = 2;

  if TOTEXP16 > 0 then has_exp = 1;
  else if TOTEXP16 = 0 then has_exp = 0;
run;
```

Stata

```
gen agecat = 1
replace agecat = 2 if agelast > 64

gen has_exp = 1
replace has_exp = 0 if (totexp16 <= 0)
```

R

```
install.packages("dplyr")
library(dplyr)

h192 = h192 %>%
  mutate(
    agecat = ifelse(AGELAST > 64, 2, 1),
    has_exp = ifelse(TOTEXP16 <= 0, 0, 1) )
```

Create new variables

Quality check on new variables

agecat	agelast		
	Min	Mean	Max
1 (< 65)	0	30.6	64
2 (65+)	65	74.0	85

has_exp	totexp16		
	Min	Mean	Max
0	0	0	0
1	1	5,407	580,640

SAS

```
proc means
proc freq
```

Stata

```
bys
sum
```

R

```
group_by
summarise
```

Process

1. Load datasets
2. Create new variables (if needed)
3. **Run survey procedure**
4. Examine results

Run survey procedure

SAS

```
proc surveymeans data = h192 mean ;
        stratum VARSTR;
        cluster VARPSU;
        weight PERWT16F;
        var TOTEXP16;
        domain has_exp * AGECAT;
run;
```

R

```
library(survey); options(survey.lonely.psu='adjust');

mepsdsgn = svydesign(
   id = ~VARPSU,
   strata = ~VARSTR,
   weights = ~PERWT16F,
   data = h192,
   nest = TRUE)

svyby(~TOTEXP16,
      by = ~agecat,
      FUN = svymean,
      design = subset(mepsdsgn, has_exp==1))
```

Stata

```
svyset [pweight=perwt16f], strata(varstr) psu(varpsu) vce(linearized) singleunit(missing)

svy, subpop(if has_exp==1): mean totexp16, over(agecat)
```

Run survey procedure

has_exp	agecat	totexp16	
		Mean	Std. Err.
1	1 (< 65)	4,615	137.6
	2 (65+)	11,781	429.2

Why survey procedures?

Correct Analysis

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	4,615	137.6
	2 (65+)	11,781	429.2

Why survey procedures?

Correct Analysis

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	4,615	137.6
	2 (65+)	11,781	429.2

Ignoring VARSTR, VARPSU

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	4,615	**122.7**
	2 (65+)	11,781	**393.1**

Why survey procedures?

Correct Analysis

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	4,615	137.6
	2 (65+)	11,781	429.2

Ignoring VARSTR, VARPSU

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	4,615	**122.7**
	2 (65+)	11,781	**393.1**

Ignoring VARSTR, VARPSU, and PERWT

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	**4,334**	**91.4**
	2 (65+)	**10,943**	**273.4**

Process

1. Load datasets
2. Create new variables (if needed)
3. Run survey procedure
4. **Examine results**

Examine results

Does output make sense?

- Population estimates

Consistent with other published results?

- Stat briefs
- Summary tables
- MEPSnet

Are estimates reliable?

- Sample size (n > 60)
- Standard errors (RSE < 0.3)

Programming checklist

- Well-defined question
- Checked the documentation
- Reserve codes addressed (-1, -9, etc.)
- Datasets merged correctly
- Adequate sample size / precision
- Survey procedures
  - PERWT, VARSTR, VARPSU
  - Using correct weights (PERWT / FAMWT / LONGWT)
  - 'domain' analysis for subsets
- Results make sense

Exercises (⭐ difficulty level)

SAS / Stata

1. National health care expenses ⭐
2. Purchases and expenses for narcotic analgesics ⭐⭐
3. Pooling multiple years of MEPS data ⭐⭐⭐
4. Pooling longitudinal files ⭐⭐⭐

https://github.com/HHS-AHRQ/MEPS-workshop

Questions?

emily.mitchell@ahrq.hhs.gov