

The George Washington University
STAT 6197 (Spring 2019): Take-Home Assignment 2 – Description

This is a description of your take-home assignment 2. For this assignment, you will use the 2010-2015 National Health Interview Survey (NHIS) sample adult files.

Step 1: Download the NHIS zipped-ASCII data files (sample adult files only) from the following web sites.

2010 Data Release https://www.cdc.gov/nchs/nhis/nhis_2010_data_release.htm

2011 Data Release https://www.cdc.gov/nchs/nhis/nhis_2011_data_release.htm

2012 Data Release (http://www.cdc.gov/nchs/nhis/nhis_2012_data_release.htm)

2013 Data Release (http://www.cdc.gov/nchs/nhis/nhis_2013_data_release.htm)

2014 Data Release (http://www.cdc.gov/nchs/nhis/nhis_2014_data_release.htm)

2015 Data Release (https://www.cdc.gov/nchs/nhis/nhis_2015_data_release.htm)

Save the downloaded-zipped files as adults2010.zip, adults2011.zip, adults2012.zip, adults2013.zip, adults2014.zip, and adults2015.zip, respectively in your computer folder. Note that the member name (i.e., samadult.dat) in six zipped files is the same.

Step 2: Read the following data fields into SAS; the column positions for these fields are specified in the SAS program that is stored in the web site. Create six separate permanent SAS data sets, one for each of the six zipped files.

Description	Data Fields	New Variables to be Created
Identifier	RECTYPE, SRVY_YR, HHX, FMX, and FPX	Create a unique identifier (ID) for each individual person
Design and analysis Variables	STRAT_P, PSU_P, WTFA_SA	None
Age	AGE_P	Age groups (18-49, 50-64, and 65+)
Sex	Sex	None
Race	RACERPI2	Set code 5 (not released) to missing.
Serious Psychological Distress (SPD)	<p>(2010-2012) SAD, NERVOUS, RESTLESS, HOPELESS, EFFORT, WORTHLS</p> <p>(2013-2015) ASISAD, ASINERV, ASIRSTLS, ASIHOPLS, ASIEFFRT, ASIWTHLS</p>	<p>1 - ALL of the time (recode as 4) 2 - MOST of the time (recode as 3) 3 - SOME of the time (recode as 2) 4 - A LITTLE of the time (recode as 1) 5 - NONE of the time (recode as 0) 7 - Refused (recode as missing) 8 - Not ascertained (recode as missing) 9 - Don't know (recode as missing)</p> <p>Create a new variable called K6SUM by taking the sum of the recoded values as specified above.</p> <p>Based on the K6SUM variable, create a grouping variable called "Serous psychological distress" (SPD) as follows: SPD = 1 IF K6SUM 13+ SPD = 0 IF K6SUM 0-12 SPD = . (MISSING). IF K6SUM=(MISSING)</p>

Step 3: Combine all six SAS datasets into a single permanent SAS dataset, and create new variables as needed (e.g., ID, 3-category age variable, race recode, and a 1/0 SPD variable) in the same DATA step.

Step 4: Use PROC FORMAT to create a permanent format catalog with entries for age group, sex, and race.

Step 5: Use PROC SURVEYMEANS to produce estimates (percentages, and standard error) by specifying the SPD variable (1/0) in the VAR statement, and age group, sex, and race variables in the DOMAIN statement. In the PROC SURVEYMEANS step, you must specify the appropriate design variables (strat_p, and psu_p), and the sampling weight (wtfa_sa). For example-SAS code on this procedure, see SAS/STAT 14.2 User's Guide and Week 7 class handouts:

<http://documentation.sas.com/?docsetId=statug&docsetTarget=titlepage.htm&docsetVersion=14.2&locale=en>

Consider the following SAS steps when writing your SAS program.

1. Data Step
2. PROC FORMAT
3. PROC SURVEYMEANS
4. PROC TRANSPOSE
5. PROC PRINT (embedded in ODS statements) or PROC REPORT

When implementing the SURVEYMEANS procedure, be sure to create an output data set that would include variable names, N, weighted proportion, and standard error, etc. You need to further process this output data set using the DATA step and PROC TRANSPOSE before you use the PROC PRINT (See the sample table below). You also need to use ODS statements to direct the PROC PRINT output to an Excel destination.

For code efficiency, you are encouraged to use arrays, macro variables, and macros.

Students are asked to complete the assignment and then submit the following items to the Blackboard folder "Take_Home_Assignment2" by 11:59 pm, April 12, 2018.

SAS Program

Log file (generated by PROC PRINTTO)

Output file (generated by PROC PRINTTO)

Output file (ODS-generated Excel format)

Table (to be populated and generated programmatically)

Table 1: Percentage of adults aged 18 or older with serious psychological distress, by age sex, and race: United States
National Health Interview Surveys, 2010-2015

Demographics	2010			2011			2012			2013			2014			2015		
	N	Percent	SE	N	Percent	SE	N	Percent	SE	N	Percent	SE	N	Percent	SE	N	Percent	SE
All																		
18-49 year																		
50-64 year																		
65+ years																		
Male																		
Female																		
White																		
African America																		
Asian																		
AIAN																		
Multiple Race																		