

Appendix C

Using Common Statistical Software Applications with the NSPY Public Use Files

APPENDIX C

USING COMMON STATISTICAL SOFTWARE APPLICATIONS WITH THE NSPY PUBLIC USE FILES

This document provides instructions and sample programs for using SAS, WesVar, and SUDAAN software to analyze the NSPY Public Use File.

C.1 Using SAS to Analyze the NSPY Public Use files

C.1.1 Sample SAS

The following SAS code (Exhibit C-1) generates a weighted frequency of two variables, gender and ethnicity. A SAS data set, Round 1, is assumed to have been downloaded or created by the user. Exhibit C-2 shows the SAS log generated by running the SAS code. Exhibit C-3 shows the output generated by running the SAS code.

Exhibit C-1. SAS program editor

```
* sample_SAS.sas - sample SAS program of frequencies ;
* Replace path-to-PUF-files with your path ;
libname PC "path-to-PUF-files" ;
options fmtsearch=(PC) ;
title1 ' National Survey of Parents and Youth (NSPY)' ;
* Sample freq of gender * Ethnicity ;
proc freq data = PC.round1 (keep = Gender2 RaceEth weight) ;
tables Gender2 * RaceEth / missing norow nocol nopercnt ;
weight weight ;
run ;
```

Exhibit C-2. SAS log

NOTE: SAS (r) Proprietary Software Release 8.2 (TS2MO)
Licensed to WESTAT INC, Site 0009005004.

NOTE: This session is executing on the WIN_PRO platform.

```
1      * sample_SAS.sas - sample SAS program of frequencies ;
2
3      * Replace path-to-PUF-files with your path ;
4      libname PC "path-to-PUF-files" ;
NOTE: Libref PC was successfully assigned as follows:
      Engine:          V8
      Physical Name:  path-to-PUF-files
5
6      options fmtsearch=(PC) ;
7
8      title1 ' National Survey of Parents and Youth (NSPY)' ;
9
10     * Sample freq of gender * Ethnicity ;
11     proc freq data = PC.round1 (keep = Gender2 RaceEth weight) ;
12       tables Gender2 * RaceEth / missing norow nocol nopercnt ;
13       weight weight ;
14     run ;
```

NOTE: There were 5411 observations read from the data set PC.ROUND1.

NOTE: PROCEDURE FREQ used:

real time	2.01 seconds
cpu time	0.34 seconds

Exhibit C-3. SAS output listing

National Survey of Parents and Youth (NSPY)				
GENDER2(YOUTH GENDER (IDENTIFIED AT ROSTER))				
RACEETH(A2(T,C), A3(T), A4(T), A5(T))				
Frequency	White or Other	African American	Hispanic	Total
Male	1.441E7	3086256	3106739	2.06E7
Female	1.357E7	3185386	2865161	1.962E7
Total	2.798E7	6271642	5971900	4.022E7

C.1.2 Sample WesVar for Variance Calculations

Since WesVar has a GUI interface, we provide a series of screen shots and commentary as well as extended output. The process illustrated started with the Round 1 youth SAS file.

Figure C-1 shows a screen shot for the WesVar interface that is used to create the WesVar file. It is for the file that was used internally at Westat to create the national effects analyses. Note that the “JKn” option is checked for “Method.” This is the correct choice for the NSPY survey and is critically important.

Figure C-2 shows a screen shot for the WesVar interface after you click “save” in the first figure, which converts the input SAS file to a “.var” file. Now all the buttons on the top become active.

Then we need to use the correct JKN factors for the replicate weights. Figure C-3 shows a screen shot for the WesVar interface after you click the button of “Attach Factors.” Figure C-4 shows a screen shot for the WesVar interface after you update the JKN factors.

Figure C-5 shows a screen shot for the WesVar interface that is used to request a table. The “RS2” and RS3” buttons request the two versions of design-based chi-square independence tests suggested by Rao and Scott, as explained in the WesVar manual.

Figure C-6 shows a screen shot for the WesVar interface for viewing a table within WesVar. This output can be exported to a text file. Attachment 1 shows that file.

Figure C-7 shows a screen shot of how the same table is displayed by WesVar TableViewer, a separate program that is provided with WesVar.

C.2 Using WesVar to Analyze the NSPY Public Use Files

To produce estimates and associated sampling variances using WesVar, it is necessary to first create a WesVar data file from either the ASCII or SAS version of the PUF dataset. Creating a WesVar data file is straightforward. Once that step is done, it is relatively easy to specify descriptive statistics such as means and proportions, tables, regression models, and other types of analyses. Simple recodes also can be performed in WesVar. Further information about the features available in WesVar are given in the WesVar User's Guide.

Below, we provide an example showing how to use WesVar (Version 4.2) to analyze NSPY data. Since WesVar uses a Graphical User Interface (GUI), we provide some simple instructions along with a series of screen shots and commentary. Examples of WesVar output are also provided. These examples use data from the Round 1 PUF.

C.2.1 Creating a WesVar file

Step 1: Double-click on the executable file **wesvar.exe** in the directory in which the WesVar software is located. The screen in Figure C-1 will appear.

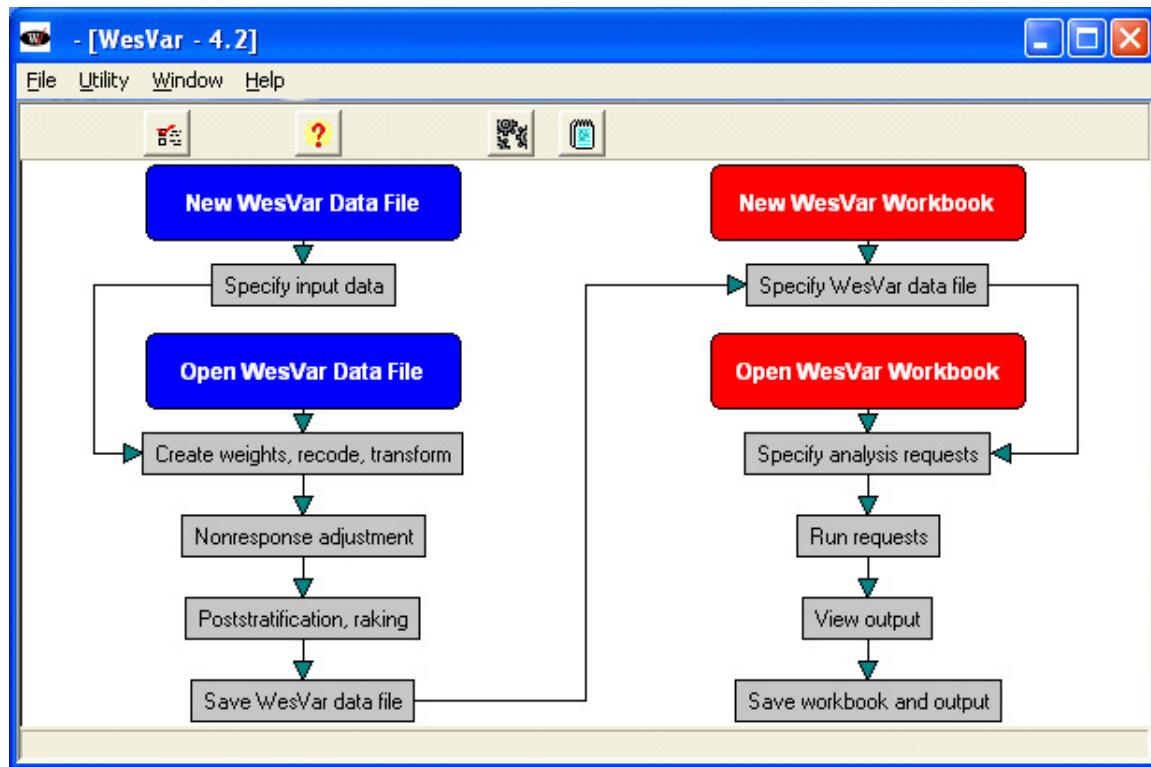


Figure C-1: Screen shot after double clicking on **wesvar.exe**

Step 2: Click on **New WesVar Data File**. The **Input Database** window will pop up. Select the appropriate directory and name of the file containing the NSPY data. To specify the file type, choose the type **SAS for Windows V7/8 (.sas7bdat)** and then click on **Open**. In our example, we use the SAS dataset: **round1.sas7bdat** as the input dataset. Another window with the name **WesVar Data File – round1.sas7bdat** will pop up (see Figure C-2). All of the variables in the SAS input file will be listed under **Source Variables**.

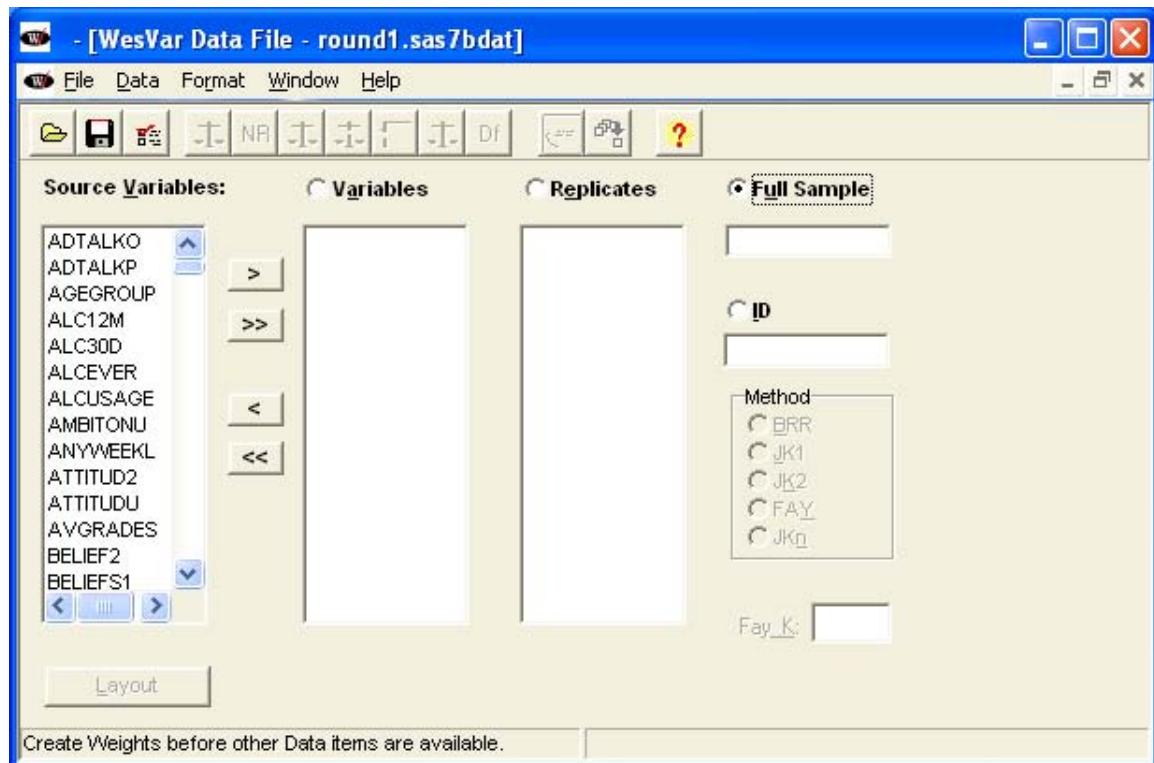


Figure C-2. Screen shot for creating a WesVar file

Step 3: Either drag or use the arrow button to move variables listed under **Source Variables** to the appropriate places shown in Figure C-3. For example, select the full-sample weight (WEIGHT) and move it to the **Full Sample** cell. Move the replicate weights (REPLW1-REPLW100) to the column under **Replicates**. Finally, select the variables to be included in the analysis from the **Source Variables** column and move them to the **Variables** column. You can select all of the remaining **Source Variables** or just the ones that you will use for your analysis.

Please note that you must choose “JKn” in the **Method** box. This is the correct choice for analyses of NSPY survey data.

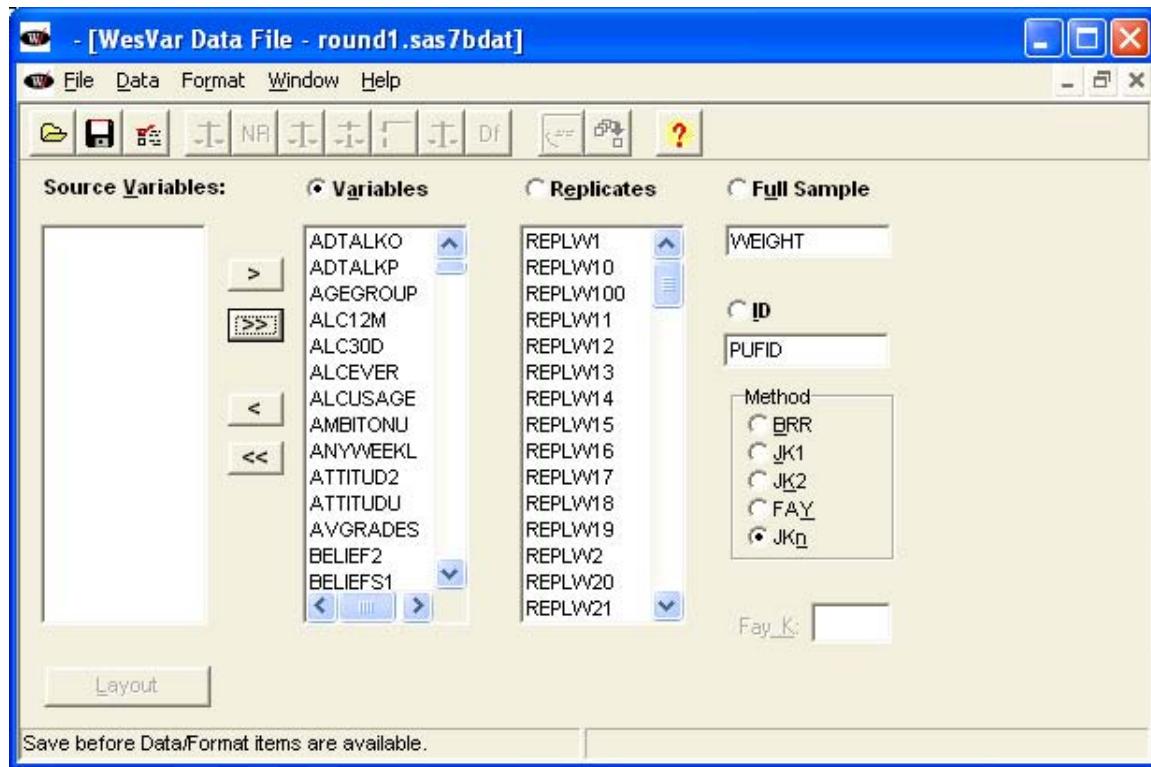


Figure C-3: Screen shot for creating a WesVar file (continued)

Step 4: Select **Save** in the **File** menu. After selecting the directory where the file is to be saved, a warning window will pop up asking if the selection of JK_n is correct. Click **Yes**. A second warning window will pop up telling you that “*The output WesVar data file shouldn’t be used for analysis until correct JK_n factors are defined using the Attach Factors Command.*” Click **OK**. The SAS input file will now be converted to WesVar format. Wait until the conversion is finished. A third window will pop up asking you to “*Make sure the degrees of freedom are consistent with the replicate method. To change the degrees of freedom, choose Modify Df.*” Click **OK**. The window shown in Figure C-3 will appear, but with all of the buttons at the top of the screen becoming active.

Step 5: Click on the **Attach Factors** button, which is to the left of the **Df** button at the top of the screen shown in Figure C-3. A new window will show up with the name **Attach Factors – round1.var** as shown in Figure C-4. Highlight the column **JK_n Factors**. Then either type in the correct JK_n factors, or click **OPEN** under **External FPC Factors** to input the JK_n factors from an existing text file.

To directly enter the required 100 JKn factors for NSPY, type in the value of 2.5677 for each of the first 60 replicates and 0.0642 for each of the remaining 40 replicates. Figure C-5 shows the screen after updating the JKn factors. When done, click **OK**. A warning message saying “*One or more JKn factors are not in range (0, 1). This may be desired for unusual replication schemes but is inconsistent with standard JKn replication. If you have a standard JKn application, you should recheck your factors. Otherwise, you may continue with your current factors.*” will pop up. Click **OK**. A **Save As** window will then appear and ask you to save the WesVar file again. Click **Save**. After the file has been saved, the screen shown in Figure C-3 will appear with all the buttons being active.

The WesVar file **round1.var** is now ready for analysis.

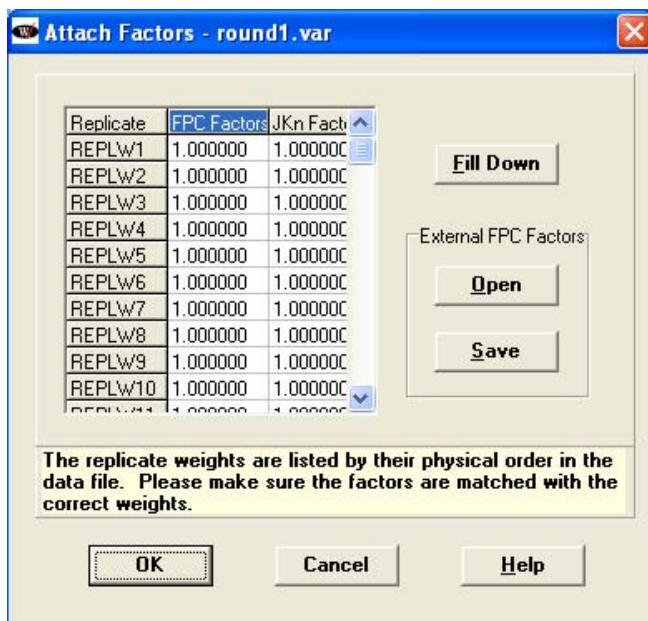


Figure C-4. Screen shot for attaching factors – Before updating the JKN Factors

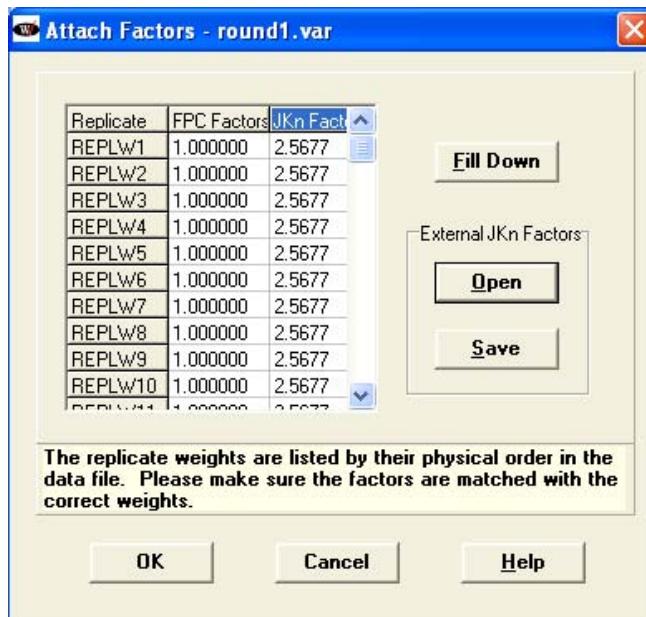


Figure C-5. Screen shot for attaching factors – After updating the JKN Factors

C.2.2 Creating a WesVar workbook

Step 1: Return to the original WesVar window (see Figure C-1) and click on **New WesVar Workbook**. A small window with the name **Open WesVar Data File for Workbook** will pop up. Select the directory and WesVar file to be used from the scrolling window and click **Open**. In our example, we choose the file **round1.var** as our input WesVar file. The WesVar workbook window shown in Figure C-6 will appear.

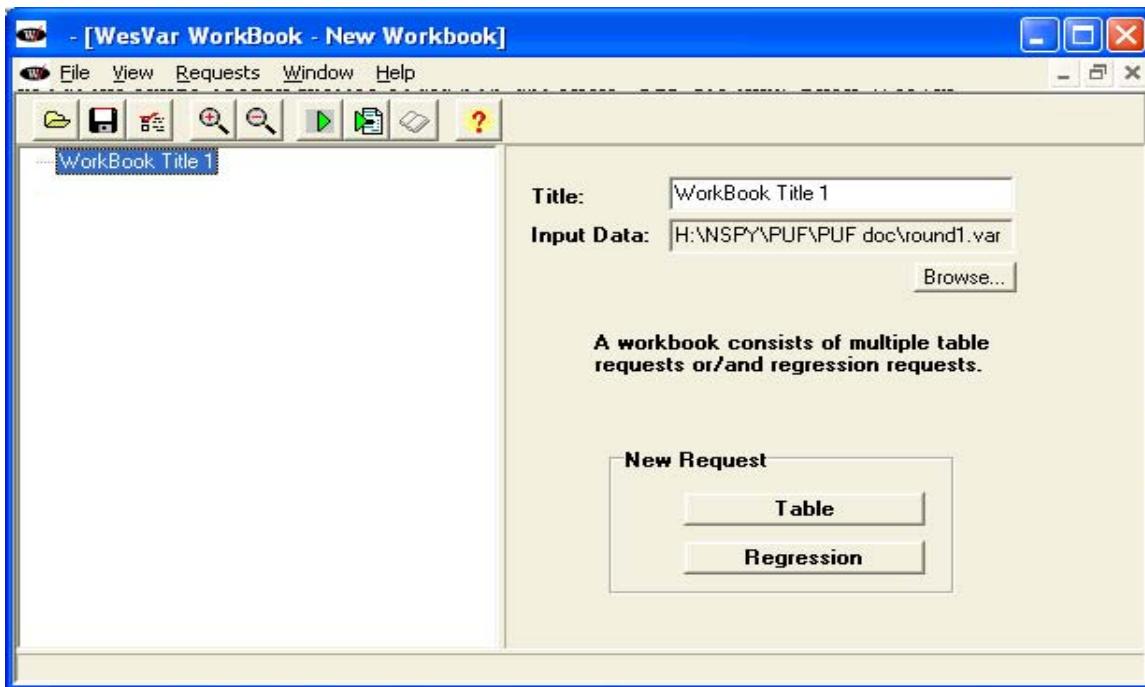


Figure C-6. Screen shot for creating a table in WesVar

Step 2: Click on **Table** under **New Request** and a new screen will show up. A variety of analyses can be specified in this workbook. For example, suppose we want to calculate (a) mean anti-marijuana belief/attitude index (MJATTBEL) and (b) proportion of youth who used marijuana in the past year (MJYEAR) by gender and race/ethnicity.

Step 3: Highlight **Table Request 1**, and then click on **Add Table Set (Single)**. **Table Set #1** will appear in the tree on the left-hand side of the screen.

Step 4: Under **Table Request 1**, highlight **Analysis Variables** and then drag MJATTBEL and MJYEAR from **Source Variables** to **Selected** on the right-hand side of the screen.

Step 5: Highlight **Compute Statistics**. On the right-hand side of the screen (See Figure C-7), type in “mean_MJATT=MEAN(MJATTBEL)” and then click on **Add as a New Entry**. Type in “mean_MJYEAR=MEAN(MJYEAR)” using the same procedure as above. Since MJYEAR is a dichotomous variable with values of 0 (meaning “no”) and 1 (meaning “yes”), the mean to be calculated here will be a proportion.

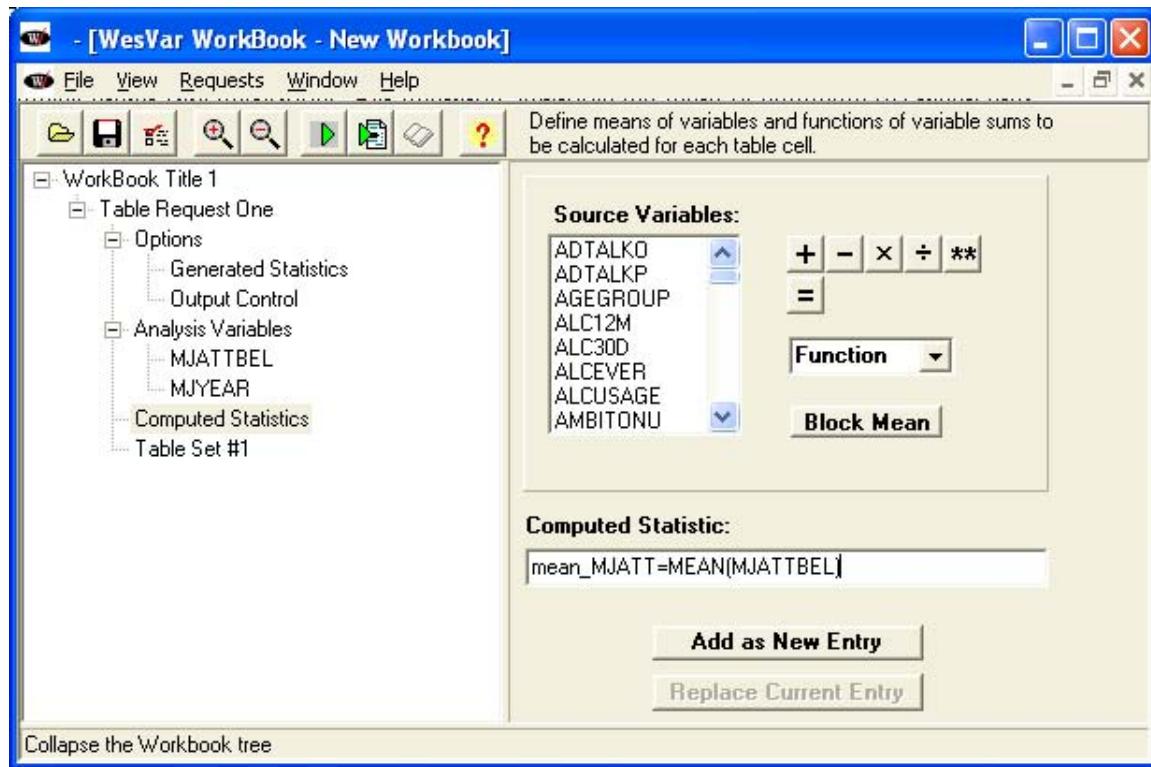


Figure C-7. Screen shot for creating a table in WesVar (continued)

Step 6: Next, highlight **Table Set #1**, and drag GENDER2 and RACEETH to the **Selected** box. Check **Value** under **Analysis Variables** on the right-hand side of the screen. Then click on **Add as a New Entry**. This step defines the table variables. Figure C-8 shows a screen shot for the workbook obtained so far.

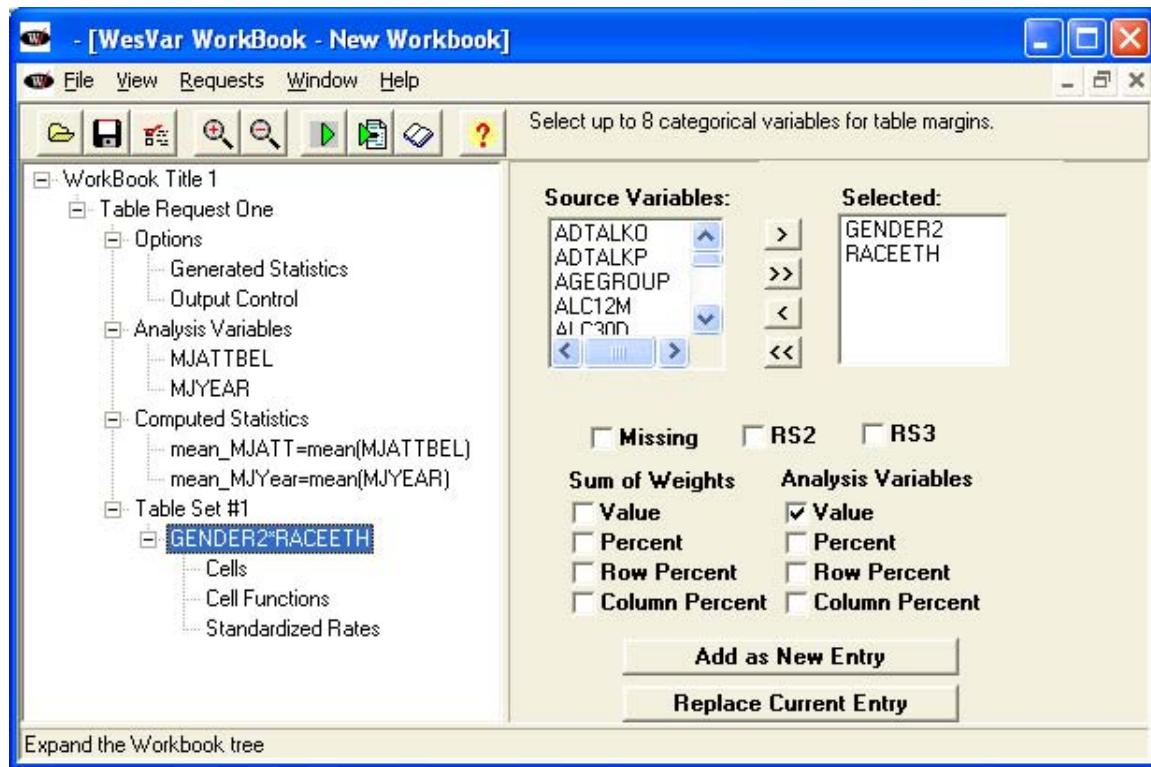


Figure C-8. Screen shot for creating a table in WesVar (continued)

Step 7: Because MJATTBEL and MJYEAR may have different numbers of missing values, we want to calculate the means of the two variables separately. To do this, highlight **Options** under **Table Request 1**, and deselect the box **Exclude all cases with missing values** on the bottom of the right-hand side of the screen.

Step 8: Save the workbook and then click on **Run Workbook Requests** at the top of the screen to create the requested table.

Step 9: To view the WesVar output, click the **View output** button at the top of the screen. Figure C-9 shows the window that will appear to view a table within WesVar. Note that for large table requests, you will need to scroll through the viewing window to see the entire table. Also note that in this example, both weighted aggregates and means are displayed. For example, all of the estimates and standard errors corresponding to the label MJYEAR in the STATISTIC column refer to estimated numbers of youth who used marijuana, while those corresponding to the label mean_MJYEAR (not displayed in Figure C-9) refer to estimated mean (proportion) of youth using marijuana. The output can

also be exported to a text file by selecting **Export** in the **File** menu. Attachment 1 (found at the end of this appendix) shows the resulting text file.

The output can also be displayed using the WesVar **TableViewer**. You will need to install the viewer first. (See pages 5-34 through 5-38 of the WesVar 4.2 User's Guide for instructions.) Figure C-10 shows how the same table is displayed by the WesVar **TableViewer**.

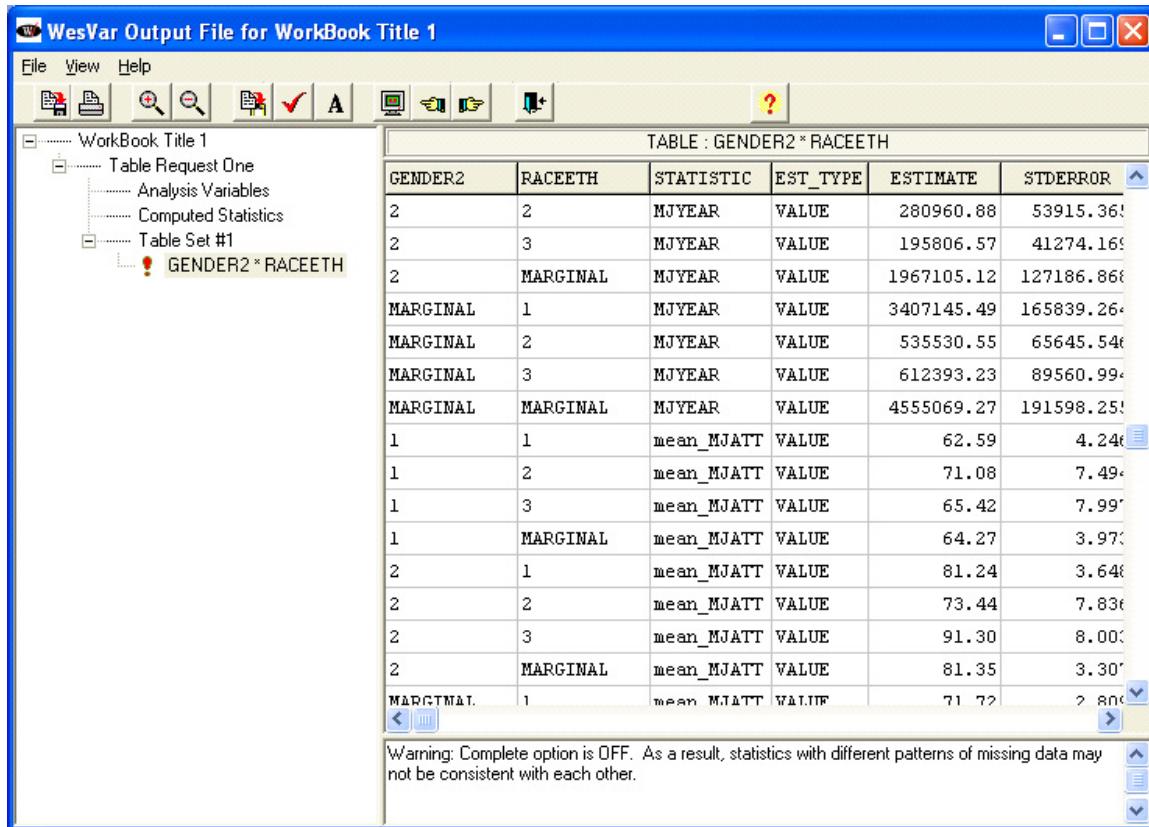


Figure C-9. Screen shot of window used to view a table in WesVar

The screenshot shows the WesVar Table Viewer interface. The title bar reads "WesVar Table Viewer - [H:\NSPY\PUF\PUF doc\round1.002]". The menu bar includes File, Edit, Window, and Help. Below the menu is a toolbar with icons for opening files, saving, and help. The main area displays a table titled "Table : GENDER2 * RACEETH" and "Variable : mean_MJATT". The table has "GENDER2" in the left column and "RACEETH" at the top. The data is presented in a grid:

	RACEETH				
	v.Estimate v.Std.Error	1	2	3	MARGINAL
1		62.59 4.246	71.08 7.494	65.42 7.997	64.27 3.973
2		81.24 3.648	73.44 7.836	91.30 8.003	81.35 3.307
MARGINAL		71.72 2.809	72.28 6.052	77.27 6.236	72.61 2.694

Figure C-10. Screen shot of table displayed by TableViewer

C.3 Using SUDAAN to Analyze the NSPY Public Use Files

Exhibit C-4 shows a sample SAS program that uses the SUDAAN DESCRIPT procedure to compute weighted estimates of (a) mean anti-marijuana belief/attitude index (MJATTBEL) and (b) proportion of youth who used marijuana in the past year (MJYEAR) by race and gender. This program starts with the same SAS file (**Round1**) used in the previous WesVar example.

Note that in the **proc descript** statement, you must specify “design = jackknife” when analyzing the NSPY PUFs. The last term in the statement is “mean” which specifies that the means of the

variables listed in the **var** statement are to be computed. The **weight** statement is used to specify the name of the full-sample weight in the NSPY dataset, and the **jackwgts** statement is used to specify the names of the corresponding replicate weights. The **jackmult** statement is used to specify the 100 JK_n factors corresponding to each replicate. Note that the required JK_n factor is 2.5677 for the first 60 replicates and 0.0642 for the last 40 replicates. Finally, the **print** statement specifies the statistics to be printed out from this run. Attachment 2 shows the output produced from this example. Note that the estimates and standard errors computed from SUDAAN are exactly the same as those computed from WesVar (see Attachment 1). Please refer to the SUDAAN User's Manual for additional details about the features and operation of SUDAAN.

Exhibit C-4. Example of SUDAAN Program

REFERENCES

Research Triangle Institute (2001). SUDAAN User's Manual, Release 8.0, Research Triangle Park, NC: Research Triangle Institute.

WesVar (2002). WesVar 4.2 User's Guide, Westat: Rockville, MD
(<http://www.westat.com/wesvar/about/wv4.2%20/manual.pdf>).

Attachments

Attachment 1. Output from WesVar Example

Summary Information of Table Request One

WESVAR VERSION NUMBER : 4.2
TIME THE JOB EXECUTED : 11:52:19 07/13/2004
INPUT DATASET NAME : H:\NSPY\PUF\PUF doc\round1.var
TIME THE INPUT DATASET CREATED : 10:30:08 07/13/2004
FULL SAMPLE WEIGHT : WEIGHT
REPLICATE WEIGHTS : REPLW1...REPLW100
VARIANCE ESTIMATION METHOD : JK_n

OPTION COMPLETE : OFF
OPTION FUNCTION LOG : ON
OPTION VARIABLE LABEL : OFF
OPTION VALUE LABEL : OFF
OPTION OUTPUT REPPLICATE ESTIMATES : OFF
FINITE POPULATION CORRECTION FACTOR : 1.00000
VALUE OF ALPHA (CONFIDENCE LEVEL %) : 0.05000 (95.00000 %)
DEGREES OF FREEDOM : 100
t VALUE : 1.984

ANALYSIS VARIABLES : MJATTBEL, MJYEAR
COMPUTED STATISTIC : mean_MJATT=mean(MJATTBEL)
mean_MJYear=mean(MJYEAR)
TABLE(S) : GENDER2*RACEETH

FACTOR(S) : 1.00
JK_n FACTOR(S) : 2.57 2.57 2.57 2.57 2.57 2.57 2.57 2.57 2.57 2.57
2.57 2.57 2.57 2.57 2.57 2.57 2.57 2.57 2.57 2.57
2.57 2.57 2.57 2.57 2.57 2.57 2.57 2.57 2.57 2.57
2.57 2.57 2.57 2.57 2.57 2.57 2.57 2.57 2.57 2.57
2.57 2.57 2.57 2.57 2.57 2.57 2.57 2.57 2.57 2.57
0.06 0.06 0.06 0.06 0.06 0.06 0.06 0.06 0.06 0.06
0.06 0.06 0.06 0.06 0.06 0.06 0.06 0.06 0.06 0.06
0.06 0.06 0.06 0.06 0.06 0.06 0.06 0.06 0.06 0.06
0.06 0.06 0.06 0.06 0.06 0.06 0.06 0.06 0.06 0.06

NUMBER OF REPLICATES : 100
NUMBER OF OBSERVATIONS READ : 5411
WEIGHTED NUMBER OF OBSERVATIONS READ : 40221299.868

Attachment 1. Output from WesVar Example (continued)

TABLE : GENDER2 * RACEETH

GENDER2	RACEETH	STATISTIC	EST_TYPE	ESTIMATE	STDERROR	CV(%)	CELL_n	DEFF
1	1	MJATTBEL	VALUE	616212099.83	40693157.079	6.604	1219	1.434
1	2	MJATTBEL	VALUE	146957957.15	16706710.856	11.368	242	1.434
1	3	MJATTBEL	VALUE	140384441.88	18130959.057	12.915	263	1.433
1	MARGINAL	MJATTBEL	VALUE	903554498.86	55600994.673	6.154	1724	1.954
2	1	MJATTBEL	VALUE	767434237.94	35300480.351	4.600	1185	1.288
2	2	MJATTBEL	VALUE	157430580.41	15601212.768	9.910	272	1.373
2	3	MJATTBEL	VALUE	165565525.96	17124075.359	10.343	231	1.752
2	MARGINAL	MJATTBEL	VALUE	1.09e+09	44535877.622	4.084	1688	1.509
MARGINAL	1	MJATTBEL	VALUE	1.38e+09	54189262.064	3.916	2404	1.375
MARGINAL	2	MJATTBEL	VALUE	304388537.56	25298517.378	8.311	514	1.725
MARGINAL	3	MJATTBEL	VALUE	305949967.85	24587316.672	8.036	494	1.505
MARGINAL	MARGINAL	MJATTBEL	VALUE	1.99e+09	73850668.934	3.704	3412	1.874
1	1	MJYEAR	VALUE	1916807.82	142447.891	7.432	1930	1.645
1	2	MJYEAR	VALUE	254569.66	55910.158	21.963	395	1.747
1	3	MJYEAR	VALUE	416586.67	67764.649	16.267	413	1.717
1	MARGINAL	MJYEAR	VALUE	2587964.15	167124.871	6.458	2738	1.655
2	1	MJYEAR	VALUE	1490337.67	117376.530	7.876	1821	1.409
2	2	MJYEAR	VALUE	280960.88	53915.365	19.190	432	1.542
2	3	MJYEAR	VALUE	195806.57	41274.169	21.079	380	1.258
2	MARGINAL	MJYEAR	VALUE	1967105.12	127186.868	6.466	2633	1.239
MARGINAL	1	MJYEAR	VALUE	3407145.49	165839.264	4.867	3751	1.243
MARGINAL	2	MJYEAR	VALUE	535530.55	65645.546	12.258	827	1.173
MARGINAL	3	MJYEAR	VALUE	612393.23	89560.994	14.625	793	1.968
MARGINAL	MARGINAL	MJYEAR	VALUE	4555069.27	191598.255	4.206	5371	1.226
1	1	mean_MJATT	VALUE	62.59	4.246	6.784	1219	1.513
1	2	mean_MJATT	VALUE	71.08	7.494	10.543	242	1.233
1	3	mean_MJATT	VALUE	65.42	7.997	12.223	263	1.283
1	MARGINAL	mean_MJATT	VALUE	64.27	3.973	6.181	1724	1.971
2	1	mean_MJATT	VALUE	81.24	3.648	4.491	1185	1.227
2	2	mean_MJATT	VALUE	73.44	7.836	10.670	272	1.592
2	3	mean_MJATT	VALUE	91.30	8.003	8.766	231	1.258
2	MARGINAL	mean_MJATT	VALUE	81.35	3.307	4.065	1688	1.495
MARGINAL	1	mean_MJATT	VALUE	71.72	2.809	3.917	2404	1.376
MARGINAL	2	mean_MJATT	VALUE	72.28	6.052	8.373	514	1.751
MARGINAL	3	mean_MJATT	VALUE	77.27	6.236	8.070	494	1.518
MARGINAL	MARGINAL	mean_MJATT	VALUE	72.61	2.694	3.711	3412	1.881
1	1	mean_MJYear	VALUE	0.13	0.010	7.202	1930	1.545
1	2	mean_MJYear	VALUE	0.08	0.018	21.392	395	1.657

1	3	mean_MJYear	VALUE	0.14	0.021	15.774	413	1.615
1	MARGINAL	mean_MJYear	VALUE	0.13	0.008	6.467	2738	1.660
2	1	mean_MJYear	VALUE	0.11	0.009	7.885	1821	1.413
2	2	mean_MJYear	VALUE	0.09	0.016	18.188	432	1.386
2	3	mean_MJYear	VALUE	0.07	0.015	21.395	380	1.296
2	MARGINAL	mean_MJYear	VALUE	0.10	0.006	6.390	2633	1.211
MARGINAL	1	mean_MJYear	VALUE	0.12	0.006	4.851	3751	1.234
MARGINAL	2	mean_MJYear	VALUE	0.09	0.011	12.284	827	1.178
MARGINAL	3	mean_MJYear	VALUE	0.10	0.015	14.679	793	1.982
MARGINAL	MARGINAL	mean_MJYear	VALUE	0.11	0.005	4.195	5371	1.219

* Warning: Complete option is OFF. As a result, statistics with different patterns of missing data may not be consistent with each other.

Attachment 2. Output from SUDAAN Example

S U D A A N
Software for the Statistical Analysis of Correlated Data
Copyright Research Triangle Institute January 2003
 Release 8.0.2

Number of observations read : 5411 Weighted count : 40221300
Denominator degrees of freedom : 100

Attachment 2. Output from SUDAAN Example (continued)

Date: 09-24-2004
Time: 16:18:09

Research Triangle Institute
The DESCRIPT Procedure

Page : 2
Table : 1

Variance Estimation Method: Replicate Weight Jackknife
by: Variable, YOUTH GENDER (IDENTIFIED AT ROSTER), A2(T,C), A3(T), A4(T), A5(T) RACE/ETHNICITY.

for: Variable = B11(T,C) - B14(C) USED MARIJUANA IN PAST YEAR.

		A2(T,C), A3(T), A4(T), A5(T) RACE/ETHNICITY			
YOUTH GENDER (IDENTIFIED AT ROSTER)		Total	White or Other	African American	Hispanic
Total	Sample Size	5371	3751	827	793
	Mean	0.11	0.12	0.09	0.10
	SE Mean	0.00	0.01	0.01	0.02
Male		2738	1930	395	413
Male	Sample Size	2738	1930	395	413
	Mean	0.13	0.13	0.08	0.14
	SE Mean	0.01	0.01	0.02	0.02
Female		2633	1821	432	380
Female	Sample Size	2633	1821	432	380
	Mean	0.10	0.11	0.09	0.07
	SE Mean	0.01	0.01	0.02	0.01

Attachment 2. Output from SUDAAN Example (continued)

Date: 09-24-2004

Time: 16:18:09

Research Triangle Institute

The DESCRIPT Procedure

Page : 2

Table : 1

Variance Estimation Method: Replicate Weight Jackknife

by: Variable, YOUTH GENDER (IDENTIFIED AT ROSTER), A2(T,C), A3(T), A4(T), A5(T) RACE/ETHNICITY.

for: Variable = B11(T,C) - B14(C) USED MARIJUANA IN PAST YEAR.

		A2(T,C), A3(T), A4(T), A5(T) RACE/ETHNICITY			
YOUTH GENDER (IDENTIFIED AT ROSTER)		Total	White or Other	African American	Hispanic
Total	Sample Size	5371	3751	827	793
	Mean	0.11	0.12	0.09	0.10
	SE Mean	0.00	0.01	0.01	0.02
Male		2738	1930	395	413
	Sample Size				
	Mean	0.13	0.13	0.08	0.14
	SE Mean	0.01	0.01	0.02	0.02
Female		2633	1821	432	380
	Sample Size				
	Mean	0.10	0.11	0.09	0.07
	SE Mean	0.01	0.01	0.02	0.01