Example 1: Variance estimates for Percentages: Women. Variance estimates using SAS, SUDAAN, STATA, and WesVar for the Percentage of Women Using the Oral Contraceptive Pill by Age

Following are the programs and output for an analysis of the percentage of females interviewed in NSFG Cycle 6 using the oral contraceptive pill during the month of interview. A cross-tabulation of the use of the oral contraceptive pill by age (in six categories: 15-19, 20-24, 25-29, 30-34, and 40-44) was generated by SAS 9.1, SUDAAN 8.0.2, STATA 8.0, and WesVar 4.1. The estimates calculated are equivalent across software. Standard errors vary slightly across packages, and design effects vary more substantially.

SAS data files were converted to STATA 8.0 and SPSS formats using DBMS/COPY 8.0. Variables in upper case are original NSFG Cycle 6 variables or recodes. Variables in lower case represent variables that were recoded as part of the variance estimation program. Library and file names are generic; the user will apply names specific to his or her computing environment. Formatting and library options have been deleted since preferences will vary across user organizations.

SAS 9.1

The DATA and SET steps create a dataset for females which contains the variables to be used in the analysis: age categories ('agerx') and use of contraceptive pill ('pill').

The PROC SURVEYFREQ produces a cross-tabulation of unweighted and weighted cell counts for the variables (i.e. 'agerx' by 'pill') specified in the TABLE statement. The WEIGHT statement identifies the weight variable FINALWGT. PROC SURVEYFREQ calculates standard errors appropriate to the complex sample design identified by the STRATUM and CLUSTER statements. The specification of ROW in the TABLE statement limits the cell counts and percentages to the row and DEFF requests calculation of the design effects for the row percentages.

SAS 9.1 Program

```
data NSFG.EX1;
set  NSFG.FEMALES;
if 15 le AGER le 19 then agerx=1;
if 20 le AGER le 24 then agerx=2;
if 25 le AGER le 29 then agerx=3;
if 30 le AGER le 34 then agerx=4;
if 35 le AGER le 39 then agerx=5;
if AGER ge 40 then agerx=6;
if CONSTAT1=6 then pill=1; else pill=2;
run;

proc surveyfreq data=NSFG.EX1;
stratum SEST;
cluster SECU_R;
weight  FINALWGT;
table agerx*pill / row deff;
run;
```

Design effects are greater than 1.0 for all but one of the row proportions due to clustering in the selection and an increase in variance due to weighting. The estimated proportions are equivalent to the other software systems.

SAS 9.1 Output

The SURVEYFREQ Procedure

Data Summary

Number of Strata	84
Number of Clusters	168
Number of Observations	7643
Sum of Weights	61560714.8

Table of agerx by pill

agerx	pill	Frequency	Weighted Frequency	Std Dev of Wgt Freq	Percent	Std Err of Percent	Design Effect	Row Percent	Std Err of Row Percent
15-19	Yes	187	1633986	176138	2.6543	0.2740	2.2211	16.6155	1.4964
	No	963	8200123	308550	13.3204	0.4921	1.6029	83.3845	1.4964
	Total	1150	9834109	380244	15.9746	0.5744	1.8784	100.000	
20-24	Yes	424	3127289	338308	5.0800	0.4776	3.6146	31.7826	1.9966
	No	939	6712331	373170	10.9036	0.4710	1.7454	68.2174	1.9966
	Total	1363	9839620	621472	15.9836	0.7570	3.2615	100.000	
25-29	Yes	313	2366080	189219	3.8435	0.2729	1.5400	25.5809	1.5872
	No	983	6883314	377552	11.1813	0.5279	2.1441	74.4191	1.5872
	Total	1296	9249394	467221	15.0248	0.6057	2.1956	100.000	
30-34	Yes	275	2234545	188101	3.6298	0.2797	1.7094	21.7527	1.4772
	No	1080	8037936	396369	13.0569	0.4906	1.6203	78.2473	1.4772
	Total	1355	10272481	477661	16.6867	0.5571	1.7059	100.000	
35-39	Yes	170	1431768	140897	2.3258	0.2393	1.9257	13.1922	1.2698
	No	1100	9421336	427176	15.3041	0.6189	2.2583	86.8078	1.2698
	Total	1270	10853104	441417	17.6299	0.6615	2.3026	100.000	
40-44	Yes	98	868678	98464	1.4111	0.1540	1.3032	7.5458	0.8347
	No	1111	10643329	625810	17.2892	0.7818	3.2666	92.4542	0.8347
	Total	1209	11512007	647860	18.7002	0.7914	3.1481	100.000	
Total	Yes	1467	11662345	590372	18.9445	0.6579	2.1540		
	No	6176	49898370	1489826	81.0555	0.6579	2.1540		
	Total	7643	61560715	1873490	100.000				

SUDAAN 8.0.2

A SAS-callable version of SUDAAN 8.0.2 was used to calculate the estimates for this example. The DATA and SET steps used to create a dataset and the variables needed for this analysis ('agerx' and 'pill'), are identical to those used above in the SAS 9.1 program and are omitted.

The PROC CROSSTAB procedure produces a frequency cross-tabulation of unweighted and weighted cell counts for the analysis variables (i.e. agerx by pill) specified in the

TABLE statement. The DESIGN used in this computation is specified as WR, with replacement. By specifying the option DEFF in the CROSSTAB statement, design effects will be calculated. The NEST statement specifies the strata (SEST) and cluster (SECU_R) variables for calculating standard errors appropriate to the complex sample design. The WEIGHT statement identifies FINALWGT for estimating the weighted frequency. The specification of NSUM, WSUM, ROWPER, SEROW, and DEFFROW in the PRINT statement limits printed output to row percentages, standard errors of row percentages, and design effects for row percentages.

SUDAAN Program

```
(same recode as required in SAS 9.1)

proc sort data=NSFG.EX1;
by SEST SECU_R;
proc crosstab data=NSFG.EX1 design=wr deff;
nest SEST SECU_R;
weight FINALWGT;
subgroup agerx pill;
levels 6 2;
table agerx * pill;
print nsum wsum rowper serow deffrow;
run;
```

The estimated percentage of women using a contraceptive pill in the six age categories are identical to those calculated by SAS 9.1:

SUDAAN 8.0.2 Output

```
                              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    :   7643    Weighted count : 61560715
Denominator degrees of freedom :     84

Variance Estimation Method: Taylor Series (WR)
by: AGERX, EA-1 R ever used Birth Control Pills?.
```

AGERX		EA-1 R ever used Birth Control Pills?		
		Total	Yes	No
Total	Sample Size	7643.0000	1467.0000	6176.0000
	Weighted Size	61560714.7761	11662344.8777	49898369.8984
	Row Percent	100.0000	18.9445	81.0555
	SE Row Percent	0.0000	0.6579	0.6579
	DEFF Row Percent			
	#4	.	2.1543	2.1543
15-19	Sample Size	1150.0000	187.0000	963.0000
	Weighted Size	9834108.6926	1633985.7873	8200122.9053
	Row Percent	100.0000	16.6155	83.3845
	SE Row Percent	0.0000	1.4964	1.4964
	DEFF Row Percent			
	#4	.	1.8587	1.8587
20-24	Sample Size	1363.0000	424.0000	939.0000
	Weighted Size	9839619.5662	3127289.0363	6712330.5299
	Row Percent	100.0000	31.7826	68.2174
	SE Row Percent	0.0000	1.9966	1.9966
	DEFF Row Percent			
	#4	.	2.5061	2.5061

SUDAAN 8.0.2 Output cont.

25-29	Sample Size	1296.0000	313.0000	983.0000
	Weighted Size	9249394.2563	2366079.9438	6883314.3125
	Row Percent	100.0000	25.5809	74.4191
	SE Row Percent	0.0000	1.5872	1.5872
	DEFF Row Percent			
	#4	.	1.7150	1.7150
30-34	Sample Size	1355.0000	275.0000	1080.0000
	Weighted Size	10272481.3018	2234545.0246	8037936.2773
	Row Percent	100.0000	21.7527	78.2473
	SE Row Percent	0.0000	1.4772	1.4772
	DEFF Row Percent			
	#4	.	1.7371	1.7371
35-39	Sample Size	1270.0000	170.0000	1100.0000
	Weighted Size	10853103.9617	1431767.5693	9421336.3924
	Row Percent	100.0000	13.1922	86.8078
	SE Row Percent	0.0000	1.2698	1.2698
	DEFF Row Percent			
	#4	.	1.7881	1.7881
40-44	Sample Size	1209.0000	98.0000	1111.0000
	Weighted Size	11512006.9975	868677.5165	10643329.4810
	Row Percent	100.0000	7.5458	92.4542
	SE Row Percent	0.0000	0.8347	0.8347
	DEFF Row Percent			
	#4	.	1.2075	1.2075

STATA 8.0

The *use* statement specifies the dataset to be used. The *svyset* command specifies the weight (FINALWGT), strata (SEST), and cluster (SECU_R) variables to be used by STATA 8.0 in estimation. These settings are saved for the current session, but can be cleared by entering the *clear* command or running *svyset* again with different settings.

The *generate* and *replace* statements create the recodes 'agerx' and 'pill'. The *svytab* command produces a cross-tabulation of 'agerx' and 'pill' and provides estimates appropriate to the complex sample design identified by the *svyset* command. The requested estimates and output are limited by specifying *row*, *deff*, and *se* after the *svytab* command.

STATA 8.0 Program

```
use "EX1.DTA"

svyset [pweight=FINALWGT], strata(SEST) psu(SECU_R)

generate agerx=1 if AGER <=19
replace  agerx=2 if AGER >=20 & AGER <=24
replace  agerx=3 if AGER >=25 & AGER <=29
replace  agerx=4 if AGER >=30 & AGER <=34
replace  agerx=5 if AGER >=35 & AGER <=39
replace  agerx=6 if AGER >=40

generate pill=2
replace pill=1 if CONSTAT1==6

svytab agerx pill, row se deff percent
```

Again, the estimated percentage of women using a contraceptive pill in the six age categories are identical to those calculated by SAS 9.1 and SUDAAN 8.0.2.

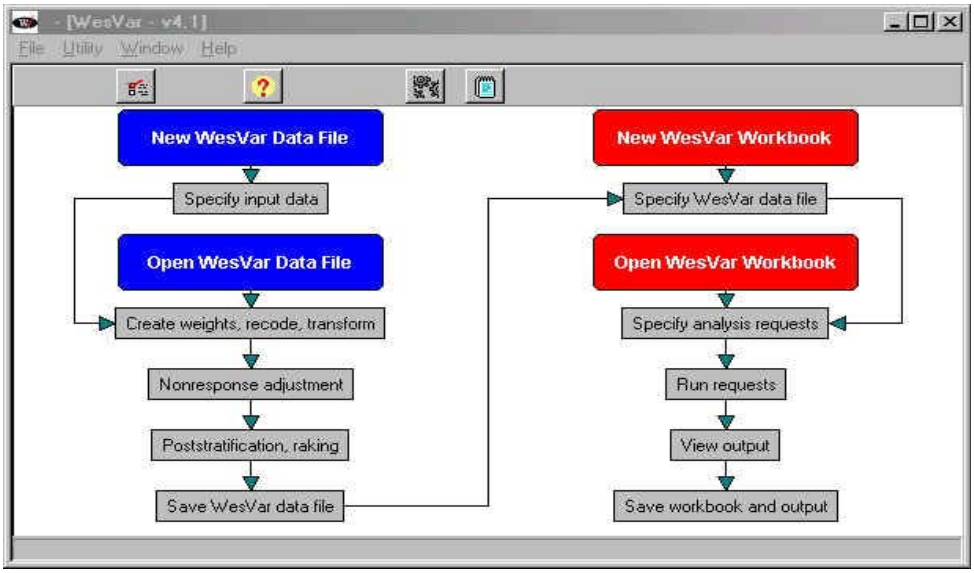
STATA 8.0 Output

```
pweight:  finalwgt                              Number of obs      =      7643
Strata:   sest                                  Number of strata   =        84
PSU:      secu_r                                Number of PSUs     =       168
                                                Population size    =  61560715


          |  EA-1 R ever used Birth
          |      Control Pills?
    agerx |     Yes       No    Total
----------+--------------------------
    15-19 |   16.62    83.38      100
          | (1.496)  (1.496)
          |   66.23    14.82
          |
    20-24 |   31.78    68.22      100
          | (1.997)  (1.997)
          |   63.18    31.36
          |
    25-29 |   25.58    74.42      100
          | (1.587)  (1.587)
          |   52.09    19.39
          |
    30-34 |   21.75    78.25      100
          | (1.477)  (1.477)
          |   47.67    14.69
          |
    35-39 |   13.19    86.81      100
          |  (1.27)   (1.27)
          |   54.24    9.506
          |
    40-44 |   7.546    92.45      100
          | (.8347)  (.8347)
          |   38.28    3.724
          |
    Total |   18.94    81.06      100
          | (.6579)  (.6579)
          |   2.154    2.154

  Key:  row percentages
        (standard errors of row percentages)
        deff for variances of row percentages

  Pearson:
    Uncorrected   chi2(5)         =  324.8924
    Design-based  F(4.63, 388.69) =   36.6663     P = 0.0000

  Mean generalized deff           =    2.0255
  CV of generalized deffs         =    0.5664
```

WesVar 4.1

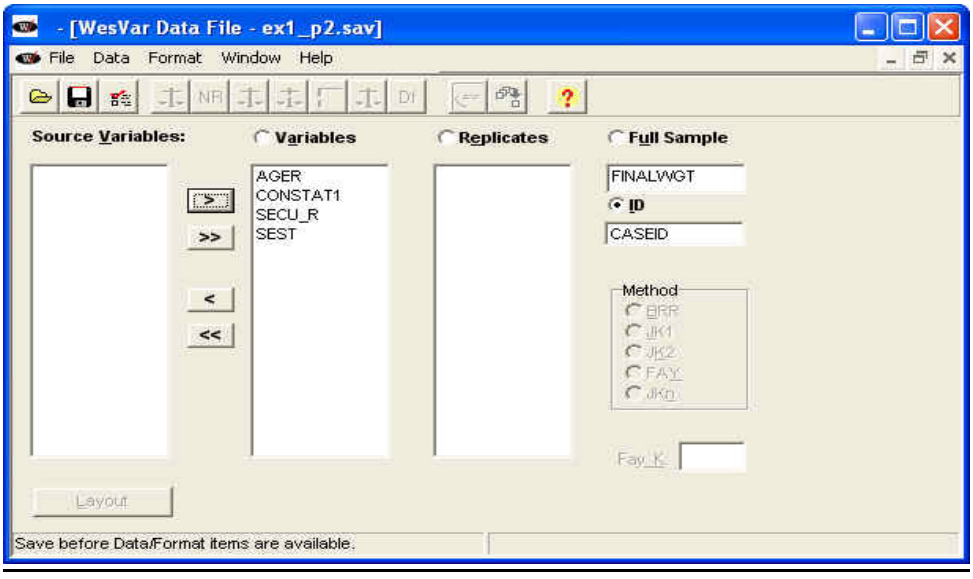
WesVar 4.1 is a windows based program. Window 1 displays the options available for initiating an analysis session. "New WesVar Data File" and the type of input file were chosen. The types of files that can be imported into WesVar 4.1 are SAS version 604, SAS transport, SPSS for windows, dBase, and ASCII files. For this example an SPSS file was imported.

WesVar 4.1 Program Window 1



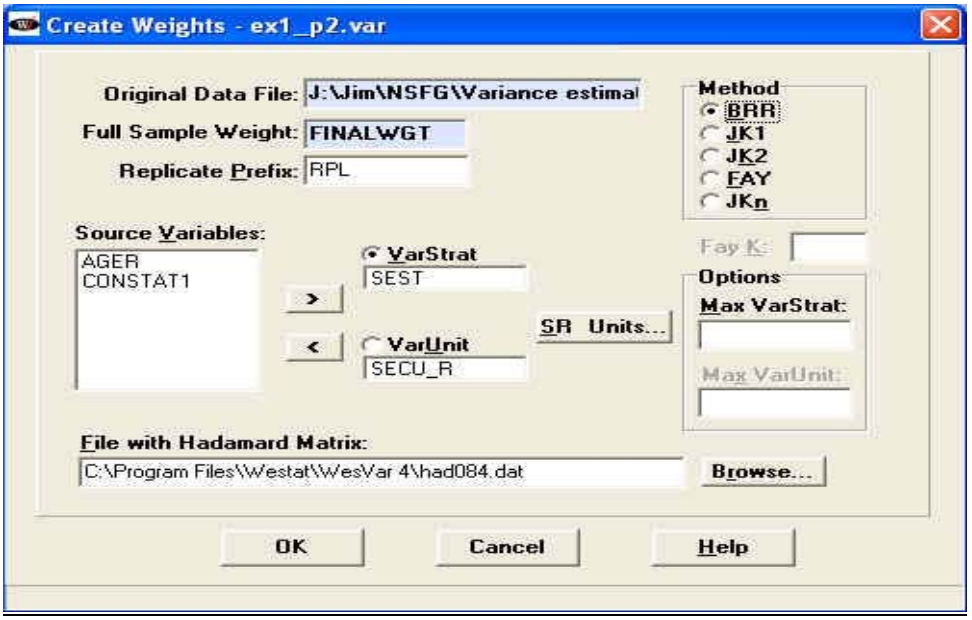
Window 2 displays the selection and categorization of variables to be used in the current analysis, the weight variable, and the sample id variable. After variables are selected and categorized, a new dataset is created.

WesVar 4.1 Program Window 2



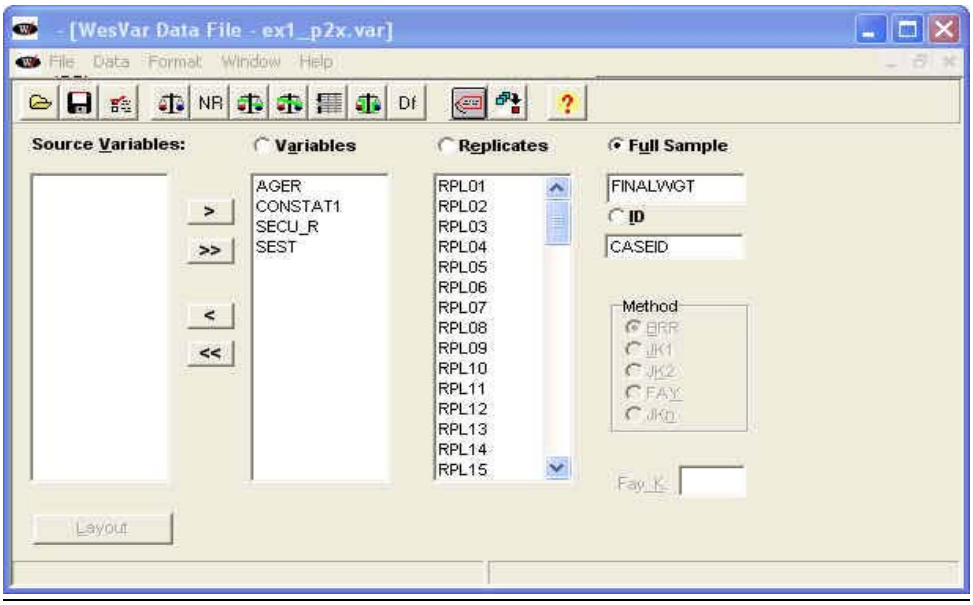
Once the dataset is saved, replicate weights are calculated by clicking on the *Create Weights* icon . In Window 3 the strata (SEST) and cluster (SECU_R) variables are specified as well as the method for estimation. In this example a balanced repeated replication method (BRR) was selected. From this window, the replicate weights are calculated and a new dataset is created.

WesVar 4.1 Program Window 3



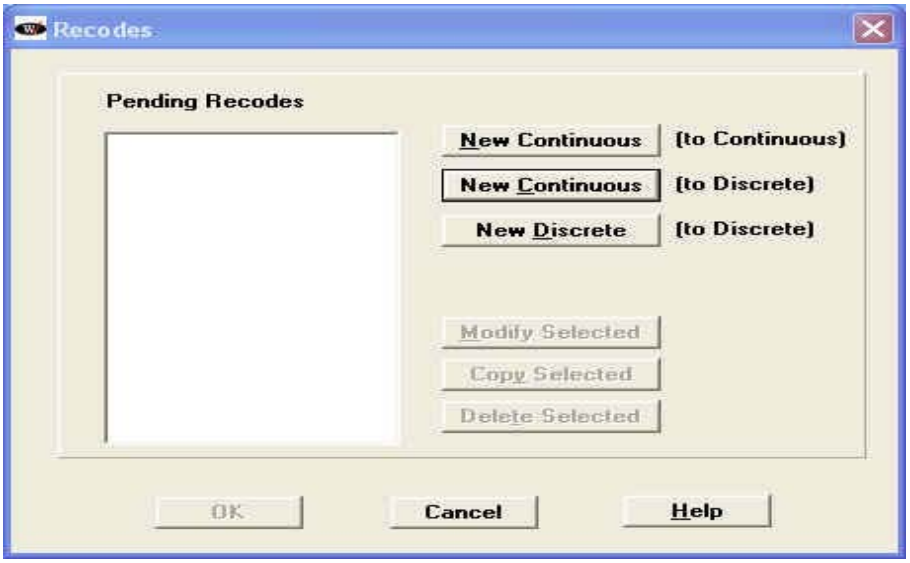
The variables on the replicate weight data file are shown in Window 4. From this window the 'agerx' recode variable was created by selecting *Recode* under the *Format* menu.

WesVar 4.1 Program Window 4

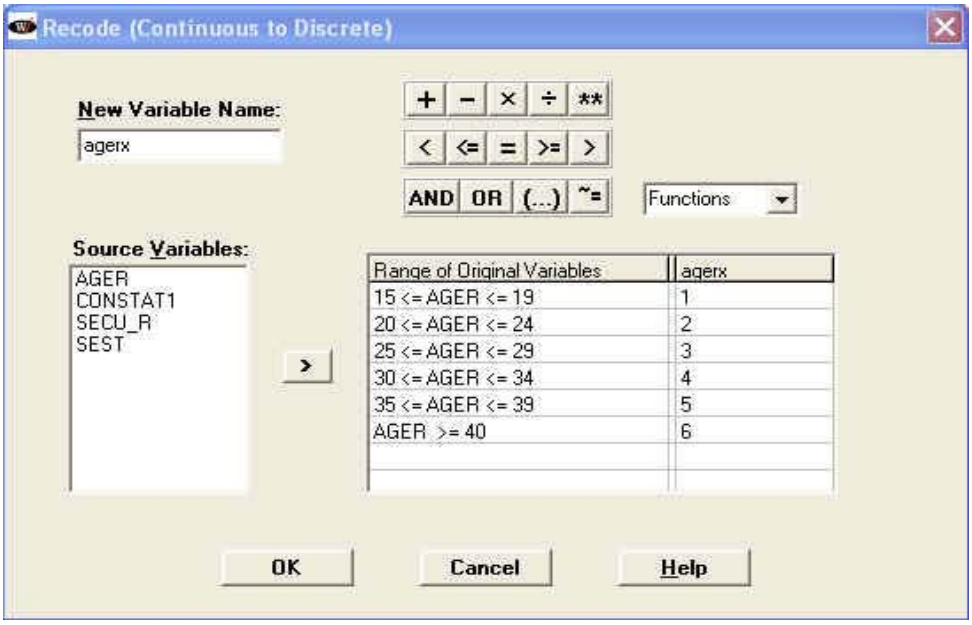


Windows 5, 6, and 7 display the procedures for recoding AGER into 'agerx' and CONSTAT1 into 'pill'. To create 'agerx' from AGER, select *New Continuous to Discrete* button; to create 'pill' from CONSTAT1, select *New Discrete to Discrete*. After the recoded variables are created, a new dataset was generated including the recodes.

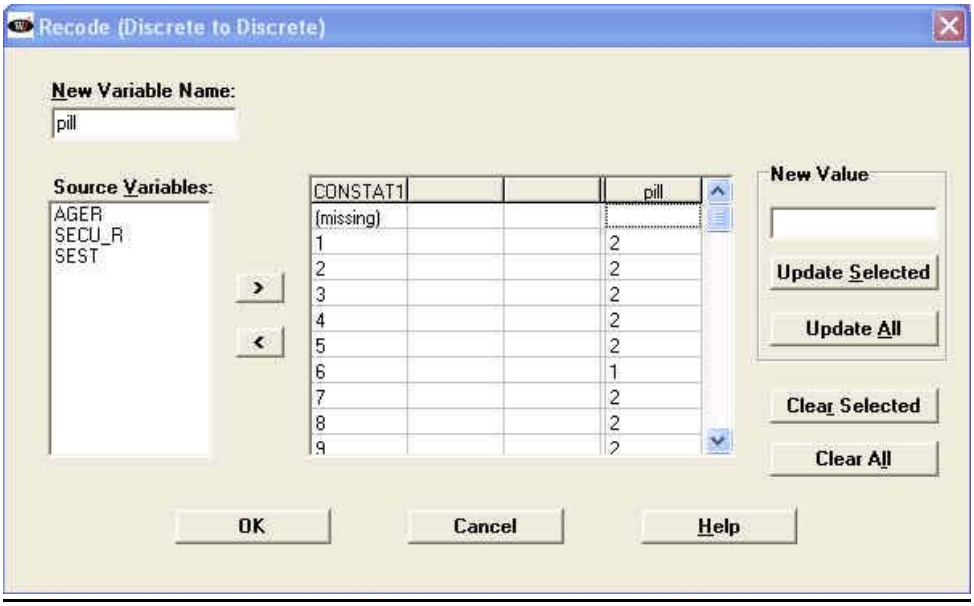
WesVar 4.1 Program Window 5



WesVar 4.1 Program Window 6

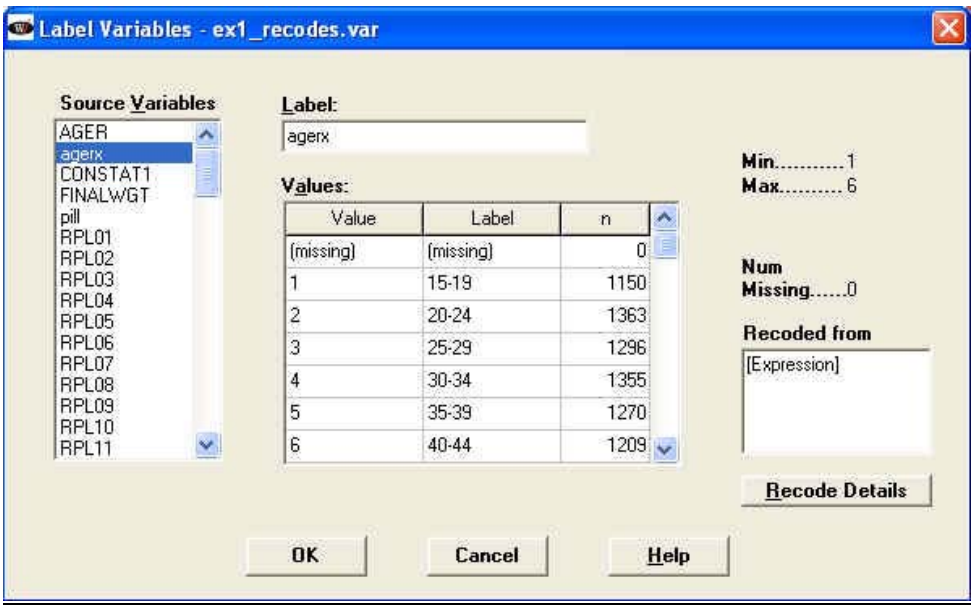


WesVar 4.1 Program Window 7

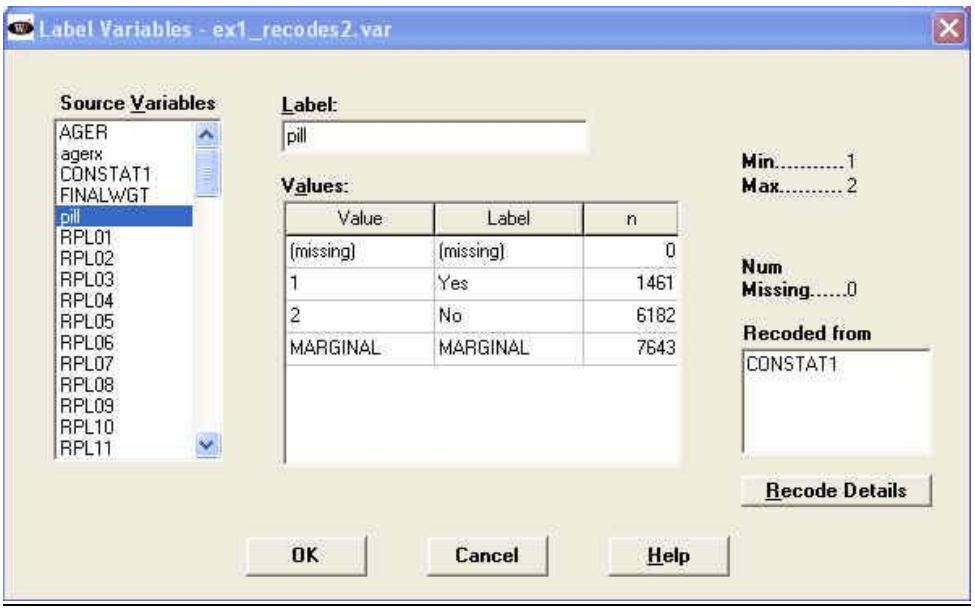


Windows 8a and 8b display how the value labels were applied. Under the *Format* menu, select *Label*. Next, select the variable ('agerx' in Window 8a) to be labeled, and enter labels in the "Label" column of the table. The process was identical for 'pill' in Window 8b.

WesVar 4.1 Program Window 8a

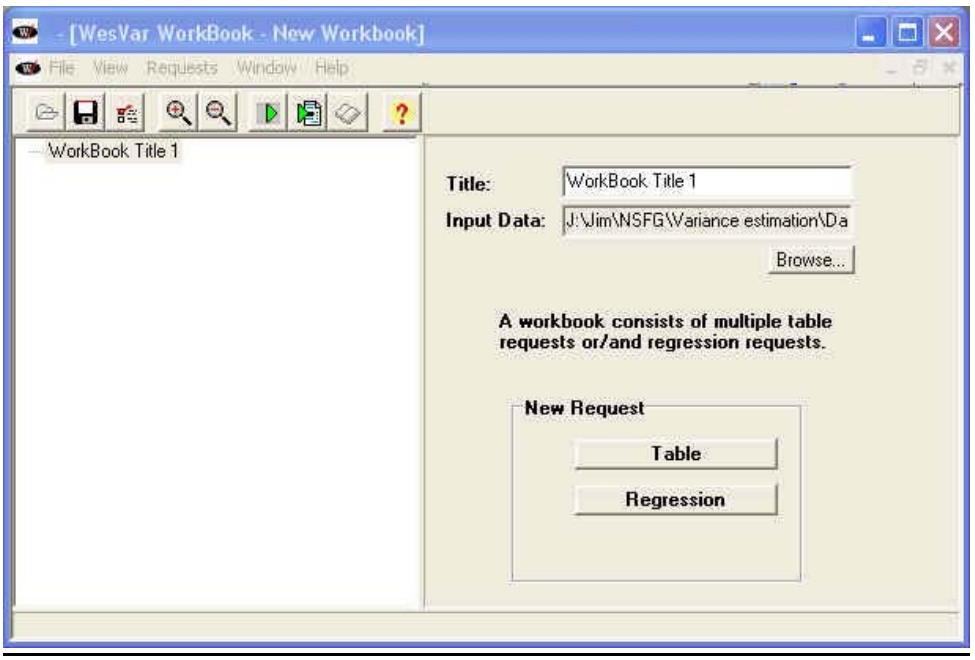


WesVar 4.1 Program Window 8b



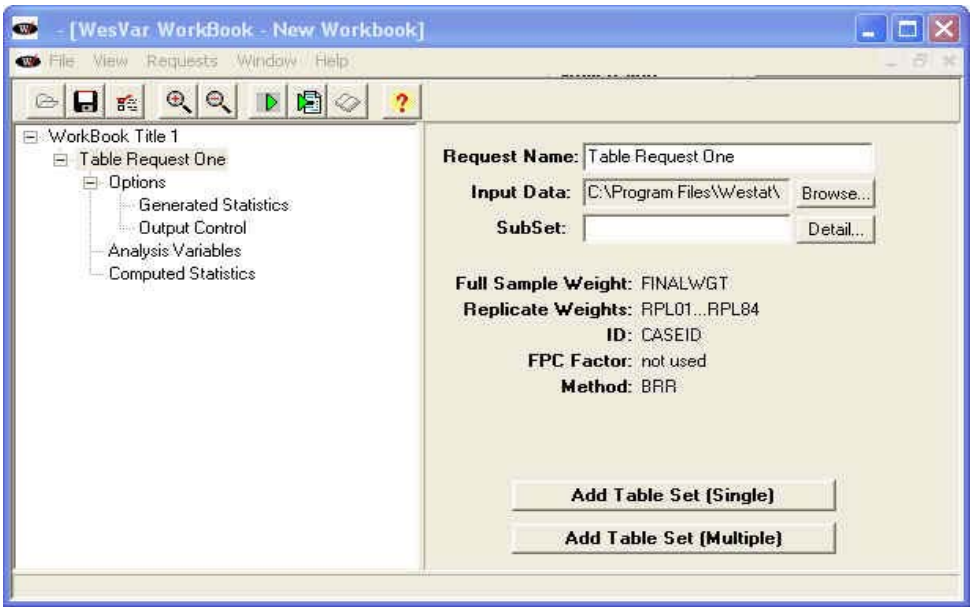
In Window 9 the type of analysis is chosen. Select *Table* button to create a cross-tabulation of 'agerx' and 'pill'.

WesVar 4.1 Program Window 9



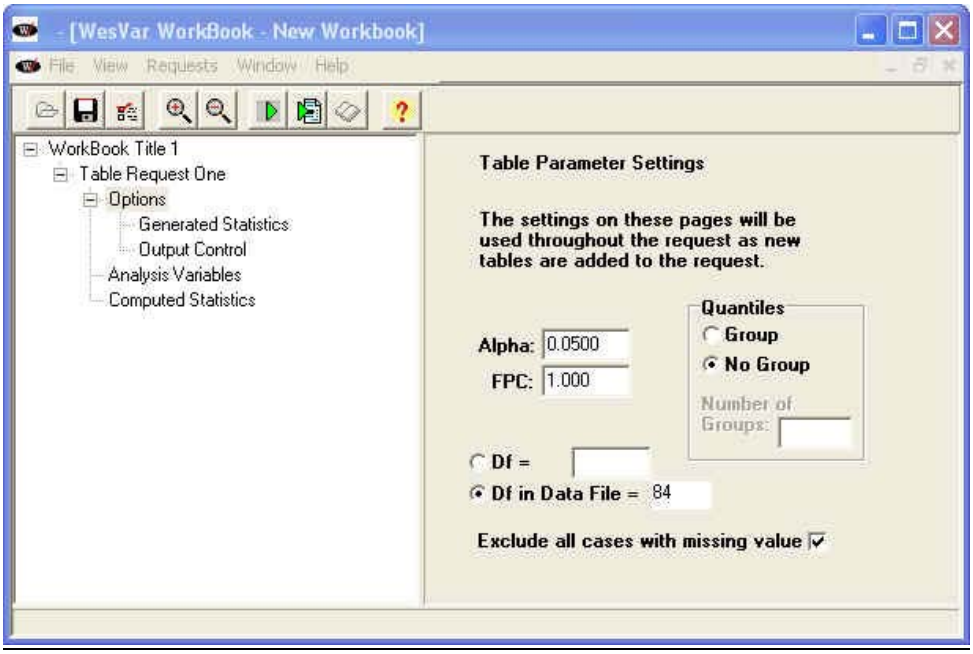
In Window 10 select *Add Table Set (Single)*.

WesVar 4.1 Program Window 10



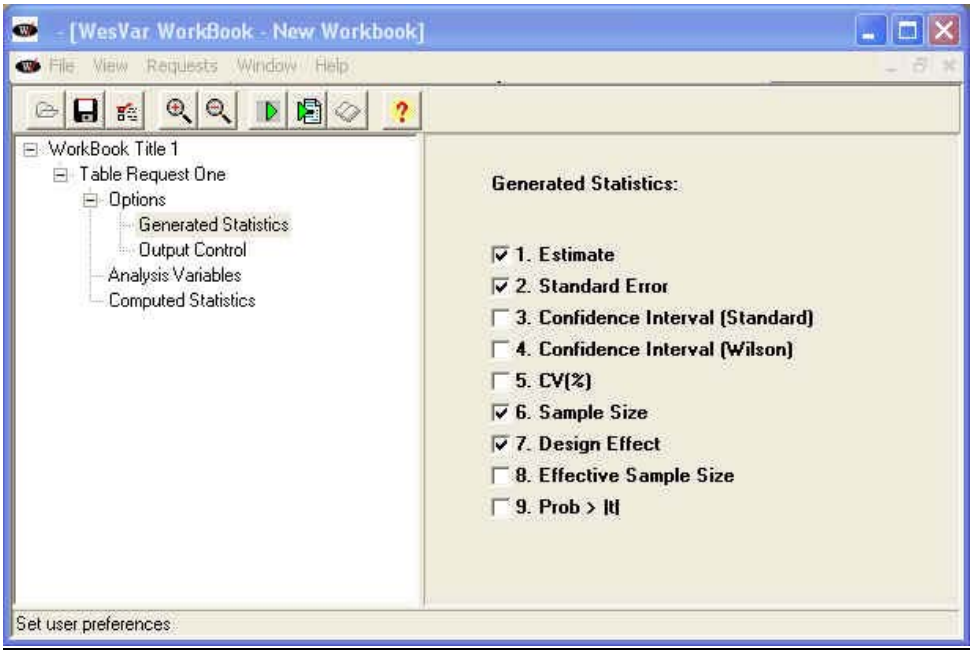
Window 11 displays the options selected for the table settings.

WesVar 4.1 Program Window 11



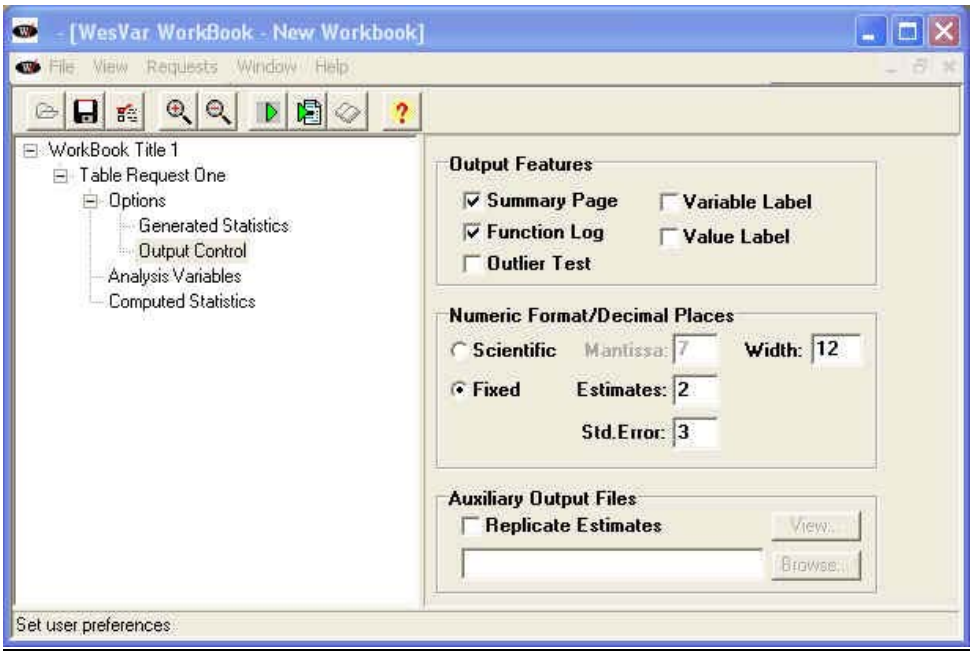
Window 12 displays the selected statistics. For this example the estimate, standard error, sample size, and design effect were chosen.

WesVar 4.1 Program Window 12



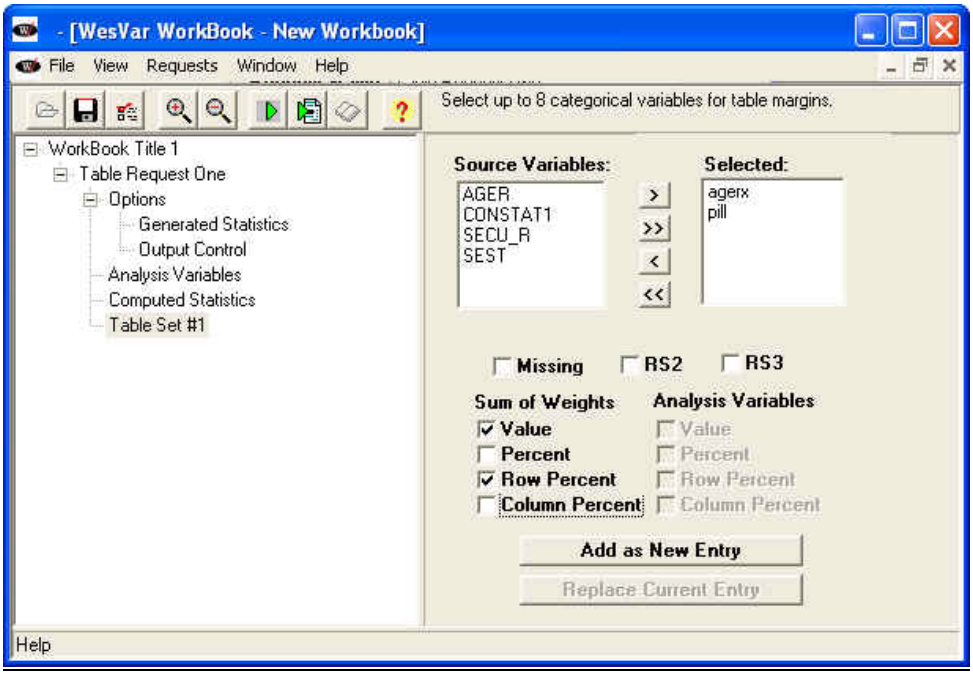
Output features selected for the table are shown in Window 13.

WesVar 4.1 Program Window 13



In Window 14 select the variables for analysis for a table estimating the percentage of women using the contraceptive pill in each of the 6 age categories. Under the *Sum of Weights* options, *Value* and *Row Percent* are selected for output.

WesVar 4.1 Program Window 14



The output provided by WesVar is a list-wise statement of all the estimates requested.

WesVar 4.1 Output

```
WESVAR VERSION NUMBER :     v4.1
TIME THE JOB EXECUTED :     12:46:43 10/06/2004
INPUT DATASET NAME :        ex1.var
TIME THE INPUT DATASET CREATED :     16:29:16 10/05/2004
FULL SAMPLE WEIGHT :        FINALWGT
REPLICATE WEIGHTS :         RPL01...RPL84
VARIANCE ESTIMATION METHOD :        BRR

OPTION COMPLETE : ON
OPTION FUNCTION LOG :       ON
OPTION VARIABLE LABEL :     OFF
OPTION VALUE LABEL :        ON
OPTION OUTPUT REPLICATE ESTIMATES :     OFF
FINITE POPULATION CORRECTION FACTOR :   1.00000
VALUE OF ALPHA (CONFIDENCE LEVEL %) :   0.05000 (95.00000 %)
DEGREES OF FREEDOM :        84
t VALUE :  1.989

ANALYSIS VARIABLES :        agerx, pill
COMPUTED STATISTIC :        None Specified.
TABLE(S) :                  agerx*pill

FACTOR(S) :         1.00

NUMBER OF REPLICATES :      84
NUMBER OF OBSERVATIONS READ :       7643
WEIGHTED NUMBER OF OBSERVATIONS READ :      61560714.776
```

WesVar 4.1 Output cont.

AGERX	PILL	STATISTIC	EST_TYPE	ESTIMATE	STDERROR	CELL_n	DENOM_n	DEFF
1	1	SUM_WTS	VALUE	1633985.79	176138.384	187	N/A	N/A
1	2	SUM_WTS	VALUE	8200122.91	308549.616	963	N/A	N/A
1	MARGINAL	SUM_WTS	VALUE	9834108.69	380244.034	1150	N/A	N/A
2	1	SUM_WTS	VALUE	3127289.04	338307.84	424	N/A	N/A
2	2	SUM_WTS	VALUE	6712330.53	373170.143	939	N/A	N/A
2	MARGINAL	SUM_WTS	VALUE	9839619.57	621472.48	1363	N/A	N/A
3	1	SUM_WTS	VALUE	2366079.94	189219.019	313	N/A	N/A
3	2	SUM_WTS	VALUE	6883314.31	377551.584	983	N/A	N/A
3	MARGINAL	SUM_WTS	VALUE	9249394.26	467221.147	1296	N/A	N/A
4	1	SUM_WTS	VALUE	2234545.02	188101.391	275	N/A	N/A
4	2	SUM_WTS	VALUE	8037936.28	396368.623	1080	N/A	N/A
4	MARGINAL	SUM_WTS	VALUE	10272481.3	477660.517	1355	N/A	N/A
5	1	SUM_WTS	VALUE	1431767.57	140897.458	170	N/A	N/A
5	2	SUM_WTS	VALUE	9421336.39	427175.751	1100	N/A	N/A
5	MARGINAL	SUM_WTS	VALUE	10853103.96	441417.214	1270	N/A	N/A
6	1	SUM_WTS	VALUE	868677.52	98463.833	98	N/A	N/A
6	2	SUM_WTS	VALUE	10643329.48	625809.971	1111	N/A	N/A
6	MARGINAL	SUM_WTS	VALUE	11512007	647860.428	1209	N/A	N/A
MARGINAL	1	SUM_WTS	VALUE	11662344.88	590371.65	1467	N/A	N/A
MARGINAL	2	SUM_WTS	VALUE	49898369.9	1489826.084	6176	N/A	N/A
MARGINAL	MARGINAL	SUM_WTS	VALUE	61560714.78	1873490.296	7643	N/A	N/A
1	1	SUM_WTS	ROWPCT	16.62	1.506	187	1150	1.883
1	2	SUM_WTS	ROWPCT	83.38	1.506	963	1150	1.883
1	MARGINAL	SUM_WTS	ROWPCT	100	.	1150	1150	.
2	1	SUM_WTS	ROWPCT	31.78	2.009	424	1363	2.536
2	2	SUM_WTS	ROWPCT	68.22	2.009	939	1363	2.536
2	MARGINAL	SUM_WTS	ROWPCT	100	.	1363	1363	.
3	1	SUM_WTS	ROWPCT	25.58	1.597	313	1296	1.736
3	2	SUM_WTS	ROWPCT	74.42	1.597	983	1296	1.736
3	MARGINAL	SUM_WTS	ROWPCT	100	.	1296	1296	.
4	1	SUM_WTS	ROWPCT	21.75	1.491	275	1355	1.769
4	2	SUM_WTS	ROWPCT	78.25	1.491	1080	1355	1.769
4	MARGINAL	SUM_WTS	ROWPCT	100	.	1355	1355	.
5	1	SUM_WTS	ROWPCT	13.19	1.273	170	1270	1.796
5	2	SUM_WTS	ROWPCT	86.81	1.273	1100	1270	1.796
5	MARGINAL	SUM_WTS	ROWPCT	100	.	1270	1270	.
6	1	SUM_WTS	ROWPCT	7.55	0.833	98	1209	1.202
6	2	SUM_WTS	ROWPCT	92.45	0.833	1111	1209	1.202
6	MARGINAL	SUM_WTS	ROWPCT	100	.	1209	1209	.
MARGINAL	1	SUM_WTS	ROWPCT	18.94	0.662	1467	7643	2.183
MARGINAL	2	SUM_WTS	ROWPCT	81.06	0.662	6176	7643	2.183
MARGINAL	MARGINAL	SUM_WTS	ROWPCT	100	.	7643	7643	.