Example 2: Variance estimates for Totals: Women. Total Number of Women Using the Oral Contraceptive Pill by Age

Following are the programs and output for estimation of the total number of females interviewed in Cycle 6 of the NSFG 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. However, due to the specific methods used in calculations, standard errors vary slightly across packages and design effects will 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 and it is assumed the user will apply names specific to his or her computing environment. Formatting and library options have been deleted; 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 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. DEFF requests calculation of the design effects for the row percentages.

SAS 9.1 Program

```
data NSFG.EX2;
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.EX2;
stratum SEST;
cluster SECU_R;
weight  FINALWGT;
table agerx*pill / 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 totals 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
15-19	Yes	187	1633986	176138	2.6543	0.2740	2.2211
	No	963	8200123	308550	13.3204	0.4921	1.6029
	Total	1150	9834109	380244	15.9746	0.5744	1.8784
20-24	Yes	424	3127289	338308	5.0800	0.4776	3.6146
	No	939	6712331	373170	10.9036	0.4710	1.7454
	Total	1363	9839620	621472	15.9836	0.7570	3.2615
25-29	Yes	313	2366080	189219	3.8435	0.2729	1.5400
	No	983	6883314	377552	11.1813	0.5279	2.1441
	Total	1296	9249394	467221	15.0248	0.6057	2.1956
30-34	Yes	275	2234545	188101	3.6298	0.2797	1.7094
	No	1080	8037936	396369	13.0569	0.4906	1.6203
	Total	1355	10272481	477661	16.6867	0.5571	1.7059
35-39	Yes	170	1431768	140897	2.3258	0.2393	1.9257
	No	1100	9421336	427176	15.3041	0.6189	2.2583
	Total	1270	10853104	441417	17.6299	0.6615	2.3026
40-44	Yes	98	868678	98464	1.4111	0.1540	1.3032
	No	1111	10643329	625810	17.2892	0.7818	3.2666
	Total	1209	11512007	647860	18.7002	0.7914	3.1481
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 for this program.

The PROC CROSSTAB procedure produces a frequency cross-tabulation of 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, SEWGT and in the PRINT statement limits printed output to sample size, weighted sample size and the standard errors of the weighted sample size.

SUDAAN Program

```
(same recode as required in SAS 9.1)
proc sort data=NSFG.EX2;
by SEST SECU_R;
proc crosstab data=NSFG.EX2 design=wr deff;
nest SEST SECU_R;
weight FINALWGT;
subgroup agerx pill;
levels 6 2;
table agerx * pill;
print nsum wsum sewgt; run;
```

The estimated total number of women in the six age categories using a contraceptive pill 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          July 2004
                                 Release 9.0.0
Variance Estimation Method: Taylor Series (WR)
by: AGERX, PILL.
```

AGERX		PILL Total	1	2
Total	Sample Size	7643.00	1467.00	6176.00
	Weighted Size	61560714.78	11662344.88	49898369.90
	Tot Percent	100.00	18.94	81.06
	SE Weighted	1873490.30	590371.65	1489826.08
	DEFF Tot Percent #4	.	2.15	2.15
1	Sample Size	1150.00	187.00	963.00
	Weighted Size	9834108.69	1633985.79	8200122.91
	Tot Percent	15.97	2.65	13.32
	SE Weighted	380244.03	176138.38	308549.62
	DEFF Tot Percent #4	1.88	2.22	1.60
2	Sample Size	1363.00	424.00	939.00
	Weighted Size	9839619.57	3127289.04	6712330.53
	Tot Percent	15.98	5.08	10.90
	SE Weighted	621472.48	338307.84	373170.14
	DEFF Tot Percent #4	3.26	3.62	1.75
3	Sample Size	1296.00	313.00	983.00
	Weighted Size	9249394.26	2366079.94	6883314.31
	Tot Percent	15.02	3.84	11.18
	SE Weighted	467221.15	189219.02	377551.58
	DEFF Tot Percent #4	2.20	1.54	2.14
4	Sample Size	1355.00	275.00	1080.00
	Weighted Size	10272481.30	2234545.02	8037936.28
	Tot Percent	16.69	3.63	13.06
	SE Weighted	477660.52	188101.39	396368.62
	DEFF Tot Percent #4	1.71	1.71	1.62
5	Sample Size	1270.00	170.00	1100.00
	Weighted Size	10853103.96	1431767.57	9421336.39
	Tot Percent	17.63	2.33	15.30
	SE Weighted	441417.21	140897.46	427175.75
	DEFF Tot Percent #4	2.30	1.93	2.26
6	Sample Size	1209.00	98.00	1111.00
	Weighted Size	11512007.00	868677.52	10643329.48
	Tot Percent	18.70	1.41	17.29
	SE Weighted	647860.43	98463.83	625809.97
	DEFF Tot Percent #4	3.15	1.30	3.27

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 "EX2.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, count se deff
```

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

STATA 8.0 Output

```
. svytab agerx pill, count se deff
pweight:  finalwgt                          Number of obs      =      7643
Strata:   sest                              Number of strata   =        84
PSU:      secu_r                            Number of PSUs     =       168
                                            Population size    =  61560715

          |                   pill
    agerx |            1              2          Total
----------+--------------------------------------------
        1 |   1633985.784    8200122.901    9834108.685
          | (176138.3837)  (308549.6153)  (380244.0336)
          |   2.421286609    1.662708252    2.172115235
          |
        2 |   3127289.038     6712330.53    9839619.568
          | (338307.8407)  (373170.1411)  (621472.4792)
          |   4.786320641    2.890570954    5.799690473
          |
        3 |   2366079.947    6883314.315    9249394.262
          | (189219.0199)  (377551.5825)  (467221.1446)
          |   1.953551435    2.894370767    3.447807708
          |
        4 |   2234545.014    8037936.278    10272481.29
          | (188101.3885)  (396368.6229)  (477660.5164)
          |   2.039649851    2.790759163    3.309425592
          |
        5 |   1431767.571    9421336.393    10853103.96
          | (140897.4575)  (427175.7508)  (441417.2131)
          |   1.762208226    2.838846146     2.70569258
          |
        6 |   868677.5125     10643329.5    11512007.01
          | (98463.83182)  (625809.9718)  (647860.4277)
          |   1.405303746    5.522661061    5.567061589
          |
    Total |   11662344.87    49898369.92    61560714.78
          | (590371.6489)  (1489826.079)
          |   4.577051322    29.14781066
```

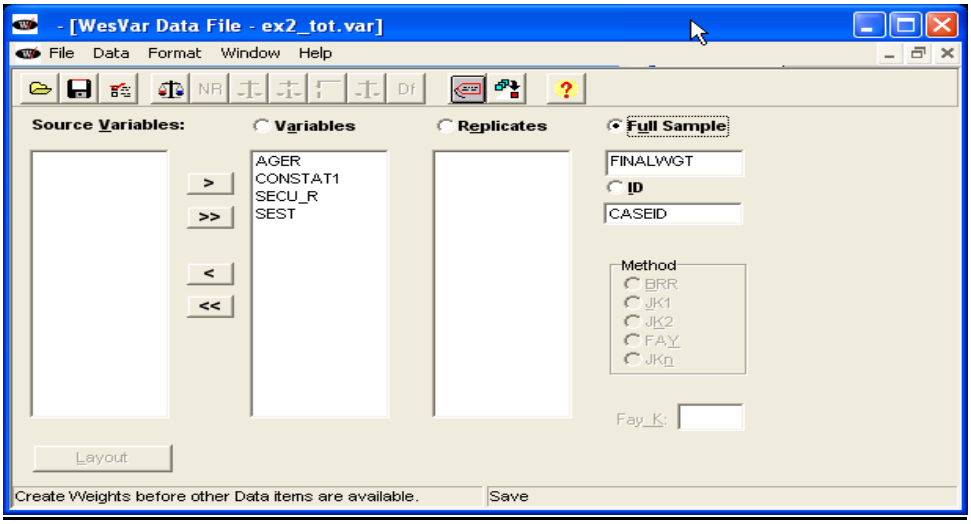
STATA 8.0 Output cont.

```
Key:  weighted counts
      (standard errors of weighted counts)
      deff for variances of weighted counts
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

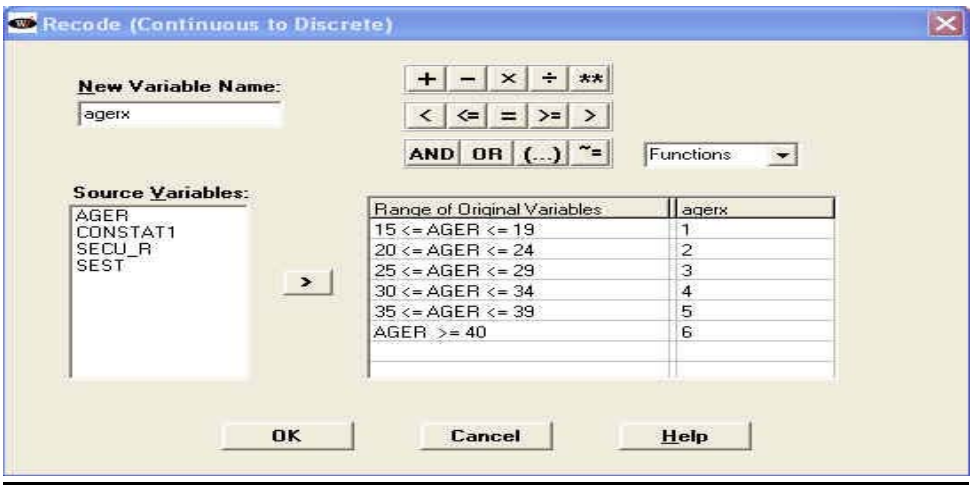
Not all WesVar windows are displayed for this example. Readers may refer to Example 1 for a full set of windows. Window 1 displays the selection and categorization of variables to be used in the current analysis.

WesVar 4.1 Program Window 1

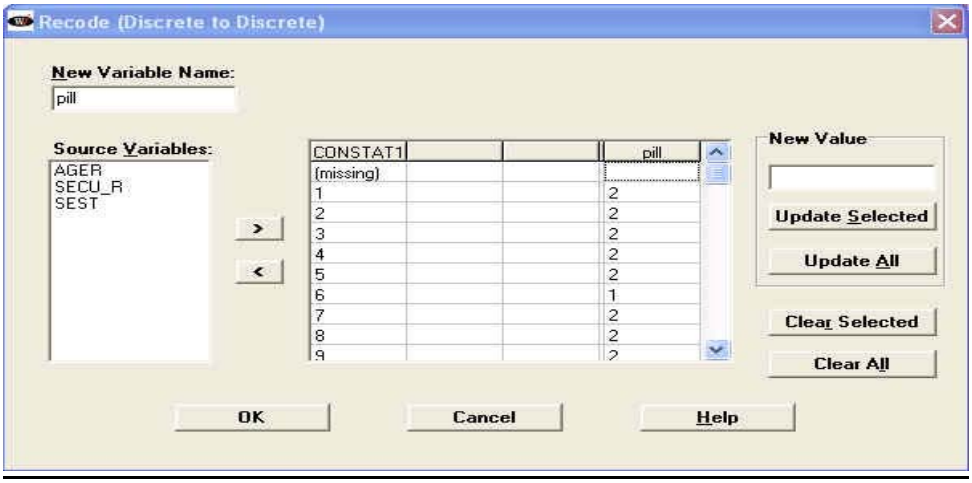


Windows 2 and 3 display the procedures for recoding AGER into 'agerx' and CONSTAT1 into 'pill'. To create 'agerx' from AGER, select *New Continuous to Discrete* button; to create 'pill' from CONSTAT1, select *New Discrete to Discrete*. After the recoded variables are created, a new dataset was generated including the recodes.

WesVar 4.1 Program Window 2

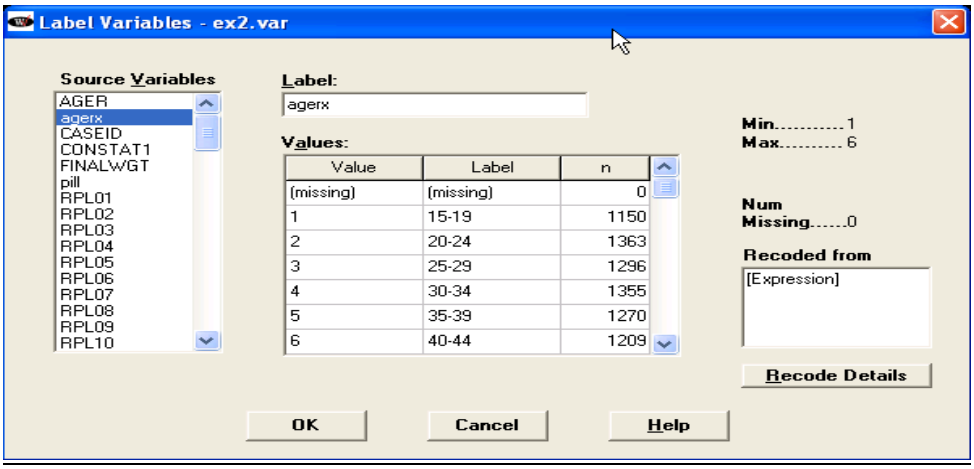


WesVar 4.1 Program Window 3

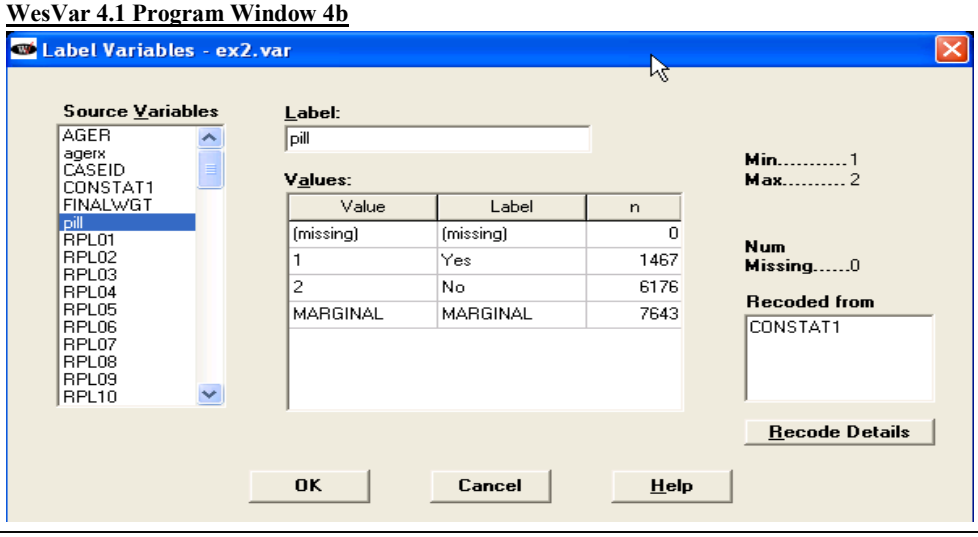


Window 4a and 4b display the how value labels were applied. Under the *Format* menu, select *Label*. Next, select the variable ('agerx' in Window 4a) to be labeled, and enter labels in the "Label" column of the table. The process was identical for 'pill' in Window 4b.

WesVar 4.1 Program Window 4a

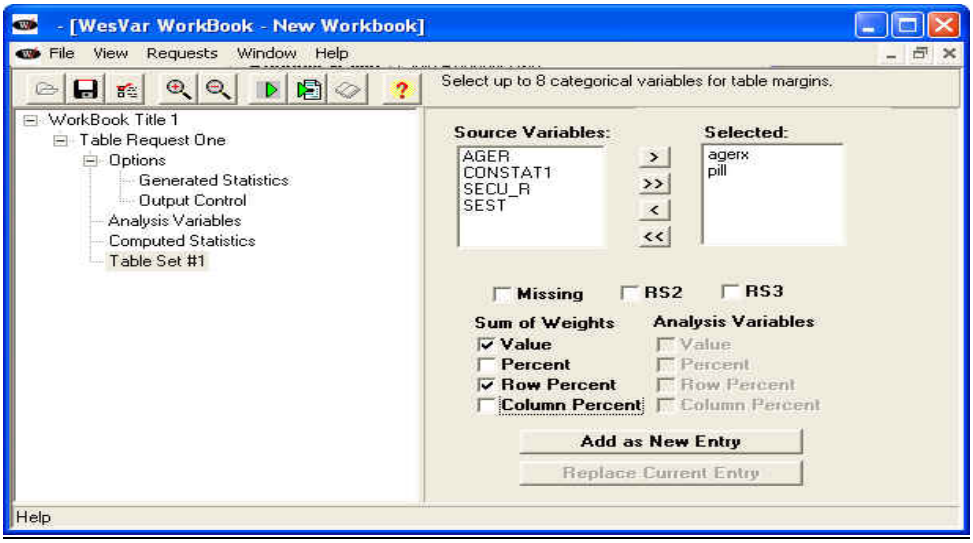


WesVar 4.1 Program Window 4b



In Window 5 select the variables for analysis for a table estimating the total number 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 5



The output provided by WesVar is a list-wise statement of all the estimates requested. WesVar does not provide design effects for totals.

WesVar 4.1 Output

```
WESVAR VERSION NUMBER :        v4.1
TIME THE JOB EXECUTED :        11:46:24 03/21/2005
INPUT DATASET NAME :           ex2.var
TIME THE INPUT DATASET CREATED :        11:44:12 03/21/2005
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
15-19	Yes	SUM_WTS	VALUE	1633985.79	176138.384	187	N/A	N/A
15-19	No	SUM_WTS	VALUE	8200122.91	308549.616	963	N/A	N/A
15-19	MARGINAL	SUM_WTS	VALUE	9834108.69	380244.034	1150	N/A	N/A
20-24	Yes	SUM_WTS	VALUE	3127289.04	338307.84	424	N/A	N/A
20-24	No	SUM_WTS	VALUE	6712330.53	373170.143	939	N/A	N/A
20-24	MARGINAL	SUM_WTS	VALUE	9839619.57	621472.48	1363	N/A	N/A
25-29	Yes	SUM_WTS	VALUE	2366079.94	189219.019	313	N/A	N/A
25-29	No	SUM_WTS	VALUE	6883314.31	377551.584	983	N/A	N/A
25-29	MARGINAL	SUM_WTS	VALUE	9249394.26	467221.147	1296	N/A	N/A
30-34	Yes	SUM_WTS	VALUE	2234545.02	188101.391	275	N/A	N/A
30-34	No	SUM_WTS	VALUE	8037936.28	396368.623	1080	N/A	N/A
30-34	MARGINAL	SUM_WTS	VALUE	10272481.3	477660.517	1355	N/A	N/A
35-39	Yes	SUM_WTS	VALUE	1431767.57	140897.458	170	N/A	N/A
35-39	No	SUM_WTS	VALUE	9421336.39	427175.751	1100	N/A	N/A
35-39	MARGINAL	SUM_WTS	VALUE	10853104	441417.214	1270	N/A	N/A
40-44	Yes	SUM_WTS	VALUE	868677.52	98463.833	98	N/A	N/A
40-44	No	SUM_WTS	VALUE	10643329.5	625809.971	1111	N/A	N/A
40-44	MARGINAL	SUM_WTS	VALUE	11512007	647860.428	1209	N/A	N/A
MARGINAL	Yes	SUM_WTS	VALUE	11662344.9	590371.65	1467	N/A	N/A
MARGINAL	No	SUM_WTS	VALUE	49898369.9	1489826.08	6176	N/A	N/A
MARGINAL	MARGINAL	SUM_WTS	VALUE	61560714.8	1873490.3	7643	N/A	N/A
15-19	Yes	SUM_WTS	ROWPCT	16.62	1.506	187	1150	1.883
15-19	No	SUM_WTS	ROWPCT	83.38	1.506	963	1150	1.883
15-19	MARGINAL	SUM_WTS	ROWPCT	100	.	1150	1150	.
20-24	Yes	SUM_WTS	ROWPCT	31.78	2.009	424	1363	2.536
20-24	No	SUM_WTS	ROWPCT	68.22	2.009	939	1363	2.536
20-24	MARGINAL	SUM_WTS	ROWPCT	100	.	1363	1363	.
25-29	Yes	SUM_WTS	ROWPCT	25.58	1.597	313	1296	1.736
25-29	No	SUM_WTS	ROWPCT	74.42	1.597	983	1296	1.736
25-29	MARGINAL	SUM_WTS	ROWPCT	100	.	1296	1296	.
30-34	Yes	SUM_WTS	ROWPCT	21.75	1.491	275	1355	1.769
30-34	No	SUM_WTS	ROWPCT	78.25	1.491	1080	1355	1.769
30-34	MARGINAL	SUM_WTS	ROWPCT	100	.	1355	1355	.
35-39	Yes	SUM_WTS	ROWPCT	13.19	1.273	170	1270	1.796
35-39	No	SUM_WTS	ROWPCT	86.81	1.273	1100	1270	1.796
35-39	MARGINAL	SUM_WTS	ROWPCT	100	.	1270	1270	.
40-44	Yes	SUM_WTS	ROWPCT	7.55	0.833	98	1209	1.202
40-44	No	SUM_WTS	ROWPCT	92.45	0.833	1111	1209	1.202
40-44	MARGINAL	SUM_WTS	ROWPCT	100	.	1209	1209	.
MARGINAL	Yes	SUM_WTS	ROWPCT	18.94	0.662	1467	7643	2.183
MARGINAL	No	SUM_WTS	ROWPCT	81.06	0.662	6176	7643	2.183
MARGINAL	MARGINAL	SUM_WTS	ROWPCT	100	.	7643	7643	.