
Example 3: Variance estimates for Means: Women. Variance estimates using SAS, SUDAAN, STATA, and WesVar for the Mean Number of Children Ever Born, by Race and Hispanic Origin and Age for Women 20-44 years of age

Following are the programs and output for an analysis of the mean number of children born to women 20-44 years of age interviewed in Cycle 6 of the NSFG, by race and Hispanic origin and age for SAS 9.1, SUDAAN 8.0.2, STATA 8.0, and WesVar 4.1. The estimates are equivalent across software. However, due to different variance estimation methods used in calculations, standard errors vary slightly.

Several conventions are utilized to display the programs. 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/her computing environment. Formatting and library options are not presented since preferences will vary across user organizations.

SAS 9.1

The DATA and SET steps create a dataset containing variables from the female dataset and the recode for age by race and hispanic origin ('group'). The PROC SURVEYMEANS step produces a table of weighted means for the variable specified in the VAR statement (PARITY). The WEIGHT statement identifies the weight variable (FINALWGT) to be used in estimating the means. PROC SURVEYMEANS calculates standard errors appropriate to the complex sample design specified in the STRATUM and CLUSTER statements. The DEFF option is not available with PROC SURVEYMEANS.

SAS 9.1 Program

```
data NSFG.EX3;
set NSFG.FEMALES;
if AGER lt 20 then delete;
if 20 le AGER le 24 then agerx=1;
if 25 le AGER le 34 then agerx=2;
if AGER ge 35 then agerx=3;
if HISPRACE=1 and agerx=1 then group=1;
if HISPRACE=1 and agerx=2 then group=2;
if HISPRACE=1 and agerx=3 then group=3;
if HISPRACE=2 and agerx=1 then group=4;
if HISPRACE=2 and agerx=2 then group=5;
if HISPRACE=2 and agerx=3 then group=6;
if HISPRACE=3 and agerx=1 then group=7;
if HISPRACE=3 and agerx=2 then group=8;
if HISPRACE=3 and agerx=3 then group=9;
if HISPRACE=4 and agerx=1 then group=10;
if HISPRACE=4 and agerx=2 then group=11;
if HISPRACE=4 and agerx=3 then group=12;
run;
proc sort data=NSFG.EX3; by group;
proc surveymeans data=NSFG.EX3;
stratum SEST;
cluster SECU_R;
weight FINALWGT;
var PARITY;
by group;
run;
```

From the output provided and as expected, design effects are large due to clustering in the design and the increase in variance due to weighting. The estimated proportions are equivalent to the other software systems.

```

SAS 9.1 Output

Mean number of children born to women 20-44 years of age
The SURVEYMEANS Procedure
Data Summary
Number of Strata          84
Number of Clusters       168
Number of Observations   6493
Sum of Weights           51726606.1

Statistics
Variable          N          Mean          Std Error          95% CL for Mean
of Mean
-----
PARITY            6493          1.502092          0.038181          1.42616568 1.57801861

Domain Analysis: group

group          Variable          N          Mean          Std Error          95% CL for Mean
of Mean
-----
Hispanic 20-24  PARITY            293          0.917116          0.089134          0.74052617 1.09370619
Hispanic 25-34  PARITY            605          1.840386          0.095041          1.65201805 2.02875331
Hispanic 35-44  PARITY            460          2.494708          0.088816          2.31873168 2.67068473
Non-Hisp White 20-24  PARITY            775          0.338462          0.029326          0.28016468 0.39676002
Non-Hisp White 25-34  PARITY           1331          1.229966          0.048924          1.13269100 1.32724049
Non-Hisp White 35-44  PARITY           1420          1.928667          0.062855          1.80367322 2.05366074
Non-Hisp Black 20-24  PARITY            218          0.795388          0.098256          0.60072506 0.99005081
Non-Hisp Black 25-34  PARITY            568          1.820761          0.072588          1.67684760 1.96467420
Non-Hisp Black 35-44  PARITY            502          2.137208          0.090084          1.95858741 2.31582897
Non-Hisp Other 20-24  PARITY             77          0.584299          0.181311          0.22548892 0.94310855
Non-Hisp Other 25-34  PARITY            147          1.083117          0.124411          0.83668391 1.32955101
Non-Hisp Other 35-44  PARITY             97          1.835584          0.269601          1.30169962 2.36946800

```

SUDAAN 8.0.2

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

The PROC DESCRIPT procedure produces descriptive statistics for the variable specified in the VAR statement (PARITY). The DESIGN used in this computation is specified as WR, with replacement. The option DEFF in the DESCRIPT statement requests that design effects be calculated. The NEST statement specifies the strata (SEST) and cluster (SECU_R) variables. The WEIGHT statement identifies FINALWGT for estimated the weighted means. SUBGROUP identifies subpopulations for which estimates are requested.

SUDAAN 8.0.2 Program

```
(same recode as required in SAS9)

proc sort data=NSFG.EX3;
by SEST SECU_R;
proc descript data=NSFG.EX3 design=wr deff;
nest SEST SECU_R;
weight FINALWGT;
var PARITY;
subgroup group;
levels 12;
run;
```

The estimated mean number of children born to women 20-44 years of age by race and Hispanic origin and age is identical to those calculated by SAS 9.1:

SUDAAN 8.0.2 Output

Mean number of children born to women 20-44 by hisrace and age

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 : 6493 Weighted count : 51726606
Denominator degrees of freedom : 84

Variance Estimation Method: Taylor Series (WR)
by: Variable, GROUP.

Variable	GROUP	Total	Hispanic 20-24	Hispanic 25-34
TOTAL NUMBER OF LIVE BIRTHS	Sample Size	6493	293	605
	Weighted Size	51726606.08	1632242.37	3248970.16
	Total	77698128.71	1496955.88	5979358.16
	Mean	1.50	0.92	1.84
	SE Mean	0.04	0.09	0.10
	DEFF Mean #4	4.71	2.11	2.90
	DEFF Total #4	8.78	4.98	8.22

Variance Estimation Method: Taylor Series (WR)
by: Variable, GROUP.

Variable	GROUP	Hispanic 35-44	Non-Hisp White 20-24	Non-Hisp White 25-34
TOTAL NUMBER OF LIVE BIRTHS	Sample Size	460	775	1331
	Weighted Size	2704983.78	6071689.25	12304528.28
	Total	6748145.22	2055038.20	15134148.32
	Mean	2.49	0.34	1.23
	SE Mean	0.09	0.03	0.05
	DEFF Mean #4	1.31	1.36	2.13
	DEFF Total #4	5.62	1.56	4.41

SUDAAN 8.0.2 Output cont.

Variance Estimation Method: Taylor Series (WR)
by: Variable, GROUP.

Variable		GROUP		
		Non-Hisp White 35-44	Non-Hisp Black 20-24	Non-Hisp Black 25-34
TOTAL NUMBER OF LIVE BIRTHS	Sample Size	1420	218	568
	Weighted Size	15788624.63	1457688.28	2694783.22
	Total	30450999.01	1159427.67	4906555.91
	Mean	1.93	0.80	1.82
	SE Mean	0.06	0.10	0.07
	DEFF Mean #4	3.10	1.60	1.24
	DEFF Total #4	12.95	1.69	4.35

Variance Estimation Method: Taylor Series (WR)
by: Variable, GROUP.

Variable		GROUP		
		Non-Hisp Black 35-44	Non-Hisp Other 20-24	Non-Hisp Other 25-34
TOTAL NUMBER OF LIVE BIRTHS	Sample Size	502	77	147
	Weighted Size	2935527.92	677999.66	1273593.90
	Total	6273834.31	396154.34	1379451.79
	Mean	2.14	0.58	1.08
	SE Mean	0.09	0.18	0.12
	DEFF Mean #4	1.61	2.77	1.63
	DEFF Total #4	6.92	4.04	3.21

Variance Estimation Method: Taylor Series (WR)
by: Variable, GROUP.

Variable		GROUP
		Non-Hisp Other 35-44
TOTAL NUMBER OF LIVE BIRTHS	Sample Size	97
	Weighted Size	935974.64
	Total	1718059.89
	Mean	1.84
	SE Mean	0.27
	DEFF Mean #4	3.52
	DEFF Total #4	8.83

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 in by STATA 8.0 in estimation. These settings are saved for the current session, but can be cleared by entering the clear command.

The *generate* and *replace* statements create the recodes *agerx* and *group*. The *svymean* command produces estimated weighted means for each of the levels of the *by* variable 'group'. The estimates provided are appropriate to the complex sample design identified

by the *svyset* command. Design effect calculations are requested by entering *deff* after the *svymean* command.

STATA 8.0 Program

```

use "EX3.dta"
drop is ager < 20
svyset [pweight=FINALWGT], strata(SEST) psu(SECU_R)
generate agerx=1 if AGER >=20 & AGER <=24
replace agerx=2 if AGER >=25 & AGER <=34
replace agerx=3 if AGER >=35

generate group=1 if HISPRACE ==1 & agerx=1
replace group=2 if HISPRACE ==1 & agerx=2
replace group=3 if HISPRACE ==1 & agerx=3
replace group=4 if HISPRACE ==2 & agerx=1
replace group=5 if HISPRACE ==2 & agerx=2
replace group=6 if HISPRACE ==2 & agerx=3
replace group=7 if HISPRACE ==3 & agerx=1
replace group=8 if HISPRACE ==3 & agerx=2
replace group=9 if HISPRACE ==3 & agerx=3
replace group=10 if HISPRACE ==4 & agerx=1
replace group=11 if HISPRACE ==4 & agerx=2
replace group=12 if HISPRACE ==4 & agerx=3

svymean PARITY, by(group) deff

```

Again, the estimated mean number of children born by race and Hispanic origin and age is identical to those calculated by SAS 9.1 and SUDAAN 8.0.2.

STATA 8.0 Output

```

. svymean parity, by(group) deff
Survey mean estimation

pweight: finalwgt          Number of obs   =      6493
Strata:   sest             Number of strata =       84
PSU:     secu_r            Number of PSUs  =      168
                               Population size = 5172606

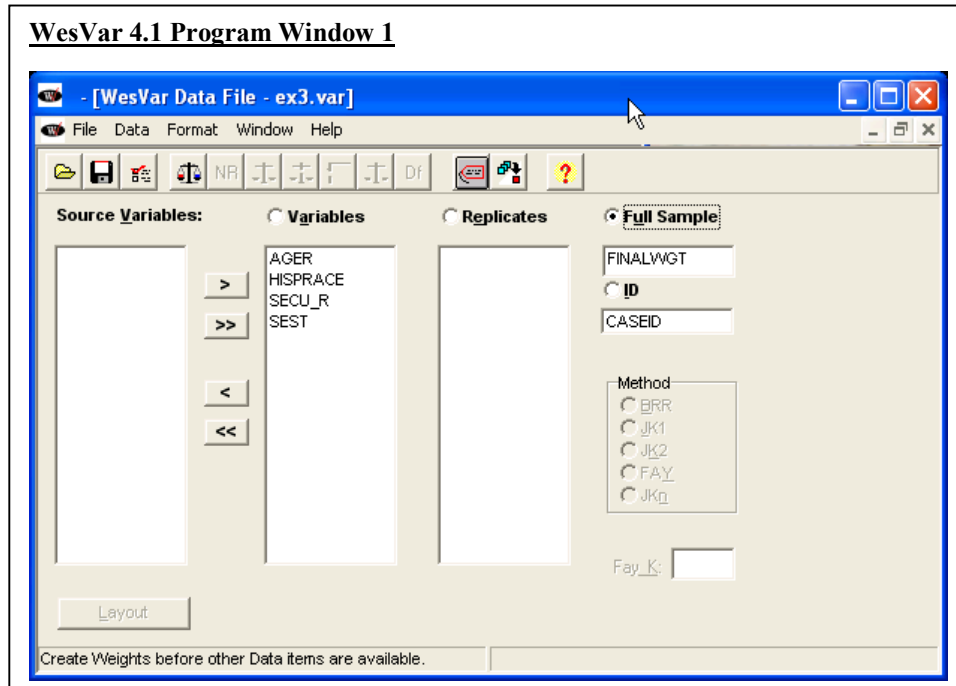
Mean   Subpop. | Estimate   Std. Err.   Deff
-----+-----+-----
parity
Hispanic 20-24 | .9171162   .0891337   1.478611
Hispanic 25-34 | 1.840386   .0950407   1.956508
Hispanic 35-44 | 2.494708   .0888155   .9661928
  White 20-24 | .3384623   .0293258   1.339018
  White 25-34 | 1.229966   .0489244   2.468895
  White 35-44 | 1.928667   .0628548   4.335046
  Black 20-24 | .7953879   .098256    1.351903
  Black 25-34 | 1.820761   .0725883   .7424528
  Black 35-44 | 2.137208   .0900844   1.187491
  Other 20-24 | .5842987   .1813113   3.105673
  Other 25-34 | 1.083117   .1244106   1.789126
  Other 35-44 | 1.835584   .2696014   4.31016

```

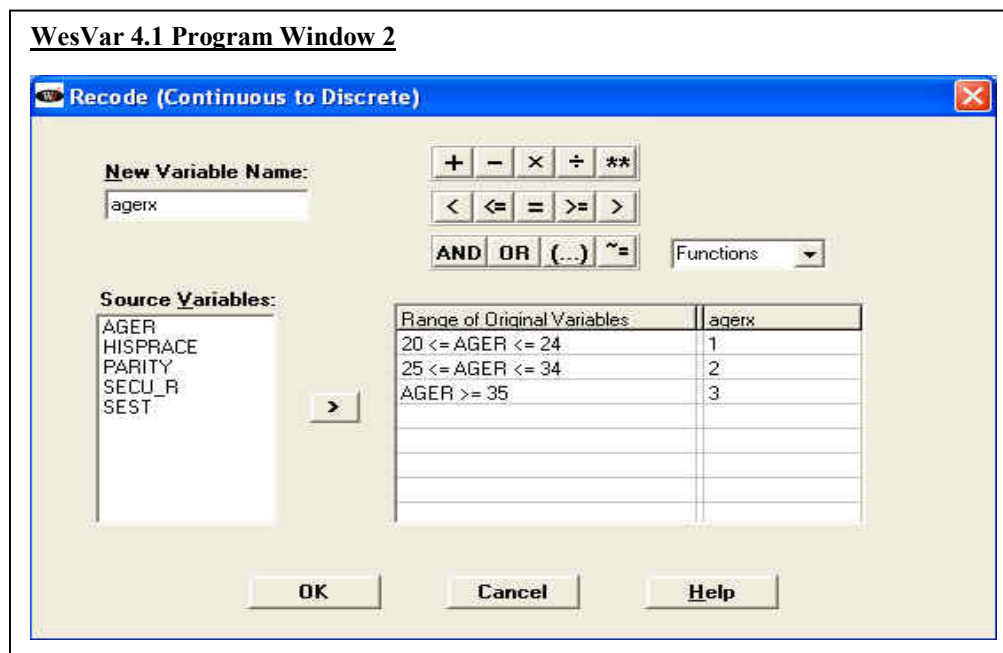
WesVar 4.1

Not all WesVar windows are displayed for this example. Readers may refer to Example 1 for the full set of windows. An SPSS file was imported for use in analysis.

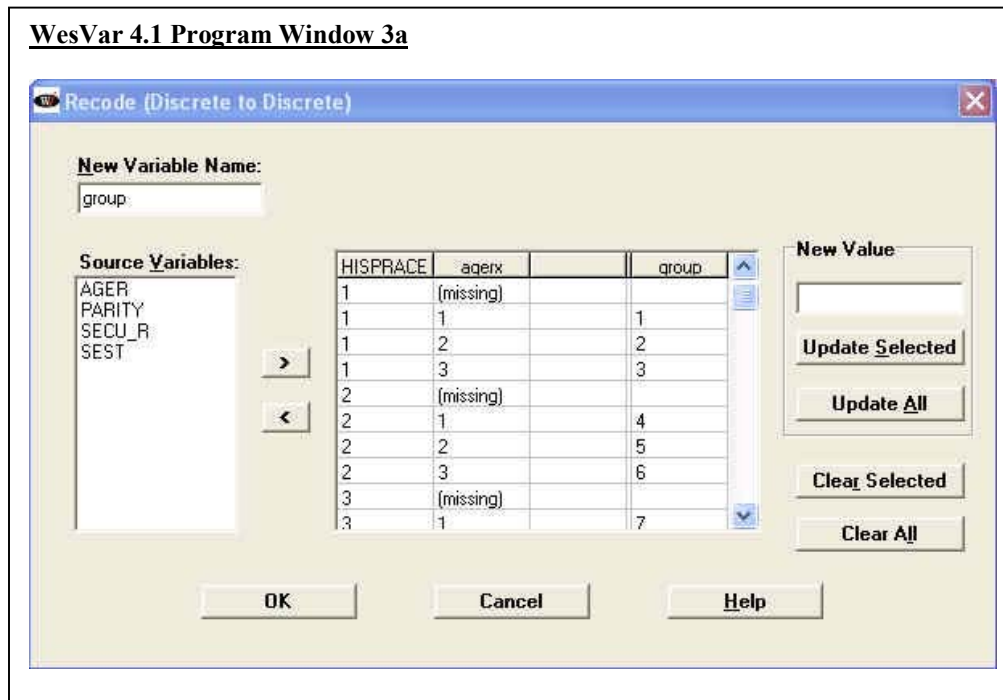
Window 1 displays the selection and categorization of variables to be used in the current analysis. After variables are selected and categorized, a new dataset is created.



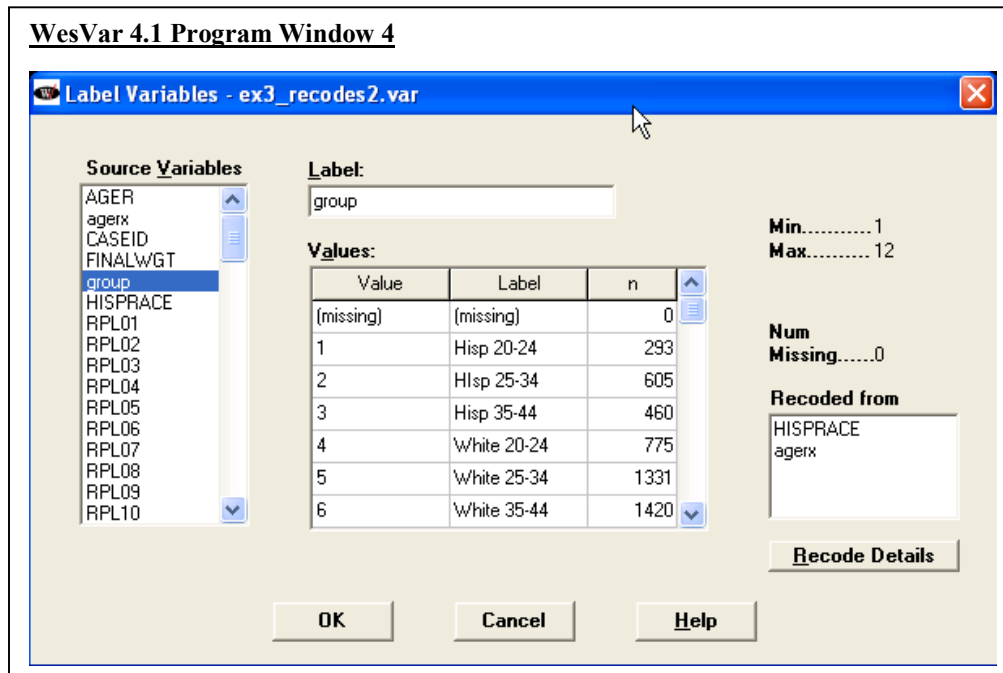
Window 2 displays the procedure for recoding AGER into 'agerx'. To create 'agerx', the *New Continuous to Discrete* button was selected.



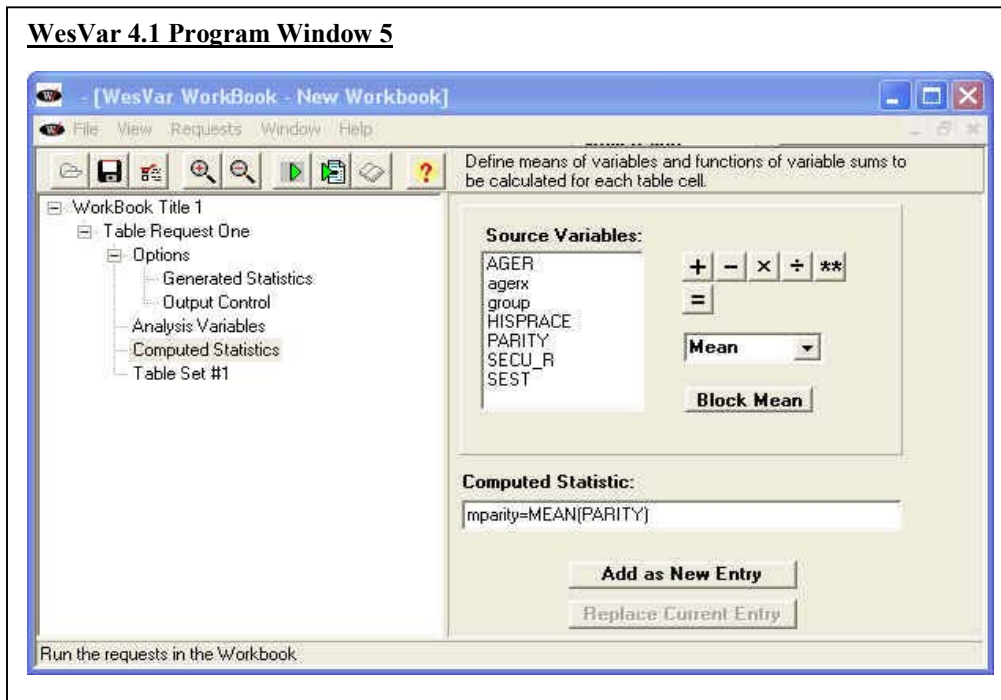
Window 3 displays the procedure for recoding HISPRACE and 'agerx' into 'group'. To create 'group', the *New Discrete to Discrete* button was selected. After the recodes are created, a new dataset is created.



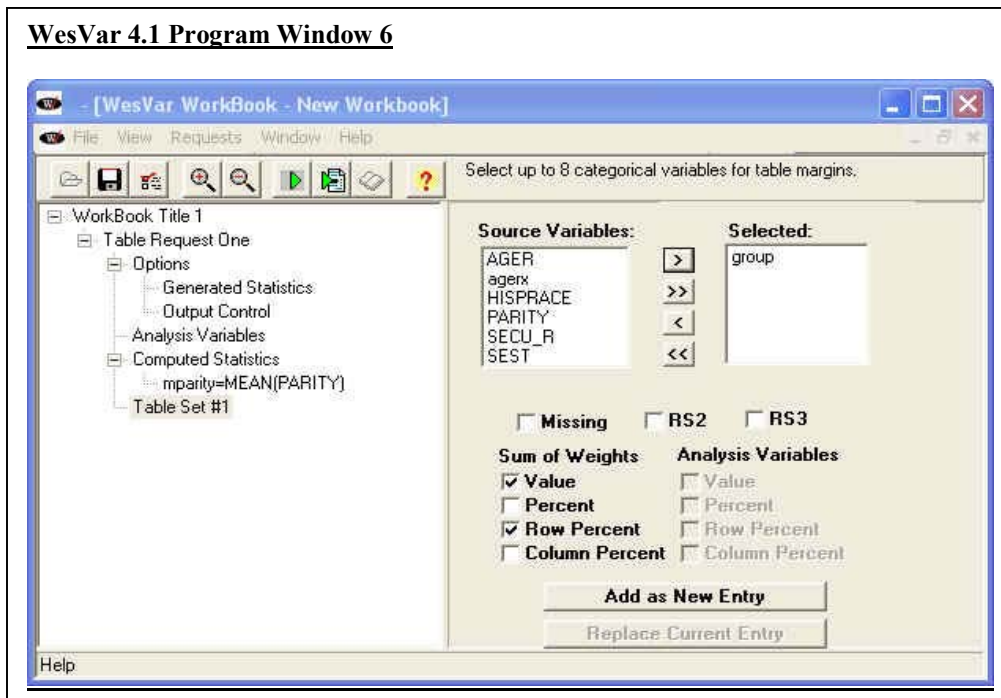
Window 4 displays how value labels were applied to 'group'. Under the *Format* menu, select *Label*.



The mean values for PARITY are calculated as shown in Window 5. To add this computation, click *Add as New Entry*.



Window 6 displays the selection of the analysis variables for a table of means by race and Hispanic origin and age. Under the *Sum of Weights, Value* and *Row Percent* are selected for output.



The output provided by WesVar is a list-wise statement of all the estimates requested. The mean number of children ever born by age and race and Hispanic origin is identical to the other software systems.

WesVar 4.1 Output

```

WESVAR VERSION NUMBER : v4.1
TIME THE JOB EXECUTED : 15:03:51 10/14/2004
INPUT DATASET NAME : ex3.var
TIME THE INPUT DATASET CREATED : 14:55:09 10/14/2004
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 : group, PARITY
COMPUTED STATISTIC : mparity=MEAN(PARITY)
TABLE(S) : group

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.

group	STATISTIC	EST_TYPE	ESTIMATE	STDERROR	CELL_n	DENOM_n	DEFF
Hispanic 20-24	SUM_WTS	VALUE	1632242.37	155303.77	293	N/A	N/A
Hispanic 25-34	SUM_WTS	VALUE	3248970.16	191938.251	605	N/A	N/A
Hispanic 35-44	SUM_WTS	VALUE	2704983.78	175452.215	460	N/A	N/A
White 20-24	SUM_WTS	VALUE	6071689.25	551372.877	775	N/A	N/A
White 25-34	SUM_WTS	VALUE	12304528.28	634524.387	1331	N/A	N/A
White 35-44	SUM_WTS	VALUE	15788624.63	810950.316	1420	N/A	N/A
Black 20-24	SUM_WTS	VALUE	1457688.28	139524.288	218	N/A	N/A
Black 25-34	SUM_WTS	VALUE	2694783.22	174944.329	568	N/A	N/A
Black 35-44	SUM_WTS	VALUE	2935527.92	228108.555	502	N/A	N/A
Other 20-24	SUM_WTS	VALUE	677999.66	102079.489	77	N/A	N/A
Other 25-34	SUM_WTS	VALUE	1273593.9	183369.153	147	N/A	N/A
Other 35-44	SUM_WTS	VALUE	935974.64	152561.558	97	N/A	N/A
MARGINAL	SUM_WTS	VALUE	51726606.08	1723792.259	6493	N/A	N/A
Hispanic 20-24	mparity	VALUE	0.92	0.092	293	N/A	2.27
Hispanic 25-34	mparity	VALUE	1.84	0.096	605	N/A	2.961
Hispanic 35-44	mparity	VALUE	2.49	0.09	460	N/A	1.339
White 20-24	mparity	VALUE	0.34	0.029	775	N/A	1.362
White 25-34	mparity	VALUE	1.23	0.05	1331	N/A	2.19
White 35-44	mparity	VALUE	1.93	0.064	1420	N/A	3.216
Black 20-24	mparity	VALUE	0.8	0.1	218	N/A	1.66
Black 25-34	mparity	VALUE	1.82	0.073	568	N/A	1.275
Black 35-44	mparity	VALUE	2.14	0.088	502	N/A	1.55
Other 20-24	mparity	VALUE	0.58	0.187	77	N/A	2.993
Other 25-34	mparity	VALUE	1.08	0.125	147	N/A	1.669
Other 35-44	mparity	VALUE	1.84	0.276	97	N/A	3.718
MARGINAL	mparity	VALUE	1.5	0.038	6493	N/A	4.764