Example 7: Variance estimates for Percentages: Men. Percentage of Males 20-44 Years of Age Who Have Ever Fathered One or More Children by Race and Hispanic Origin

Following are the programs and output for an analysis of the percentage of males interviewed in Cycle 6 of the NSFG who have ever fathered one or more children by race and Hispanic origin 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 male dataset and a recode, fathered one or more children ('biokidsx').

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

SAS 9.1 Program

```
data NSFG.EX7;
set  NSFG.MALES;
if BIOKIDS gt 0 then biokidsx=1; else biokidsx=2;
if AGER lt 20 then delete;
run;

proc surveyfreq data=NSFG.EX7;
stratum SEST;
cluster SECU;
weight  FINALWGT;
var HISPRACE*biokidsx / row deff;
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

The SURVEYFREQ Procedure

Data Summary

Number of Strata	84
Number of Clusters	168
Number of Observations	3807
Sum of Weights	50938980

Table of HISPRACE by biokidsx

HISPRACE	biokidsx	Frequency	Weighted Frequency	Std Dev of Wgt Freq	Percent	Std Err of Percent	Design Effect	Row Percent	Std Err of Row Percent
HISPANIC	1 or more children	509	5584670	366544	10.9635	0.7095	1.9626	65.2384	2.4483
	No children	379	2975728	250412	5.8418	0.4511	1.4082	34.7616	2.4483
	Total	888	8560398	438909	16.8052	0.7814	1.6621	100.000	
NON-HISPANIC WHITE	1 or more children	737	17637646	953904	34.6250	1.5435	4.0060	52.6851	2.1045
	No children	1242	15839852	1105234	31.0957	1.5286	4.1507	47.3149	2.1045
	Total	1979	33477499	1522376	65.7208	1.3367	3.0188	100.000	
NON-HISPANIC BLACK	1 or more children	378	3488507	247523	6.8484	0.5624	1.8872	60.1025	2.2939
	No children	347	2315754	196636	4.5461	0.3808	1.2715	39.8975	2.2939
	Total	725	5804261	353443	11.3945	0.7777	2.2797	100.000	
NON-HISPANIC OTHER	1 or more children	78	1653841	215889	3.2467	0.4035	1.9722	53.4045	4.2777
	No children	137	1442981	207102	2.8328	0.3687	1.8801	46.5955	4.2777
	Total	215	3096822	329851	6.0795	0.5699	2.1651	100.000	
Total	1 or more children	1702	28364665	1080896	55.6836	1.7276	4.6031		
	No children	2105	22574315	1362192	44.3164	1.7276	4.6031		
	Total	3807	50938980	1773994	100.000				

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 CROSSTAB procedure produces a cross-tabulation of unweighted and weighted cell counts for the analysis variables HISPRACE by 'biokidsx' specified in the TABLE statement. The DESIGN used in this computation is specified as WR, with replacement. The option DEFF in the CROSSTAB statement requests that design effects be calculated. The NEST statement specifies the strata (SEST) and cluster (SECU) variables. The WEIGHT statement identifies the weight variable FINALWGT. The specification of NSUM, WSUM, ROWPER, SEROW, and DEFFROW in the PRINT statement limits printed output to those quantities.

SUDAAN 8.0.2 Program

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

proc sort data=NSFG.EX7;
by SEST SECU;
proc crosstab data=NSFG.EX7 design=wr deff;
nest SEST SECU;
weight FINALWGT;
subgroup HISPRACE biokidsx;
levels 4 , 2;
table HISPRACE * biokidsx;
print nsum wsum rowper serow deffrow;
run;
```

The estimated percentage of men having fathered one or more children by race and Hispanic origin calculated by SUDAAN 8.0.2 are identical to those from 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    :   3807    Weighted count : 50938980
Denominator degrees of freedom :     84

Variance Estimation Method: Taylor Series (WR)
by: Race and hispanic origin, BIOKIDSX.
```

Race and hispanic origin		BIOKIDSX Total	1 or more children	No children
Total	Sample Size	3807.0000	1702.0000	2105.0000
	Weighted Size	50938979.9902	28364664.5720	22574315.4182
	Row Percent	100.0000	55.6836	44.3164
	SE Row Percent	0.0000	1.7276	1.7276
	DEFF Row Percent #4	.	4.6043	4.6043
HISPANIC	Sample Size	888.0000	509.0000	379.0000
	Weighted Size	8560398.1206	5584669.8799	2975728.2407
	Row Percent	100.0000	65.2384	34.7616
	SE Row Percent	0.0000	2.4483	2.4483
	DEFF Row Percent #4	.	2.3471	2.3471
NON-HISPANIC WHITE	Sample Size	1979.0000	737.0000	1242.0000
	Weighted Size	33477498.5928	17637646.3392	15839852.2536
	Row Percent	100.0000	52.6851	47.3149
	SE Row Percent	0.0000	2.1045	2.1045
	DEFF Row Percent #4	.	3.5161	3.5161
NON-HISPANIC BLACK	Sample Size	725.0000	378.0000	347.0000
	Weighted Size	5804261.2031	3488507.0282	2315754.1750
	Row Percent	100.0000	60.1025	39.8975
	SE Row Percent	0.0000	2.2939	2.2939
	DEFF Row Percent #4	.	1.5910	1.5910
NON-HISPANIC OTHER	Sample Size	215.0000	78.0000	137.0000
	Weighted Size	3096822.0736	1653841.3247	1442980.7489
	Row Percent	100.0000	53.4045	46.5955
	SE Row Percent	0.0000	4.2777	4.2777
	DEFF Row Percent #4	.	1.5810	1.5810

STATA 8.0

The *use* statement specifies the dataset to be used. The *svyset* command specifies the weight (FINALWGT), strata (SEST), and cluster (SECU) 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 recode *biokidsx*. The *svytab* command produces a cross-tabulation of HISPRACE and *biokidsx* 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 "EX7.DTA"

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

generate biokidsx=2
replace biokidsx=1 if BIOKIDS >0

drop if AGER < 20

svytab HISPRACE biokidsx, row se deff percent
```

Again, the estimated percentages of men having fathered one or more children by race and hispanic origin are identical to those calculated by SAS 9.1 and SUDAAN 8.0.2.

STATA 8.0 Output

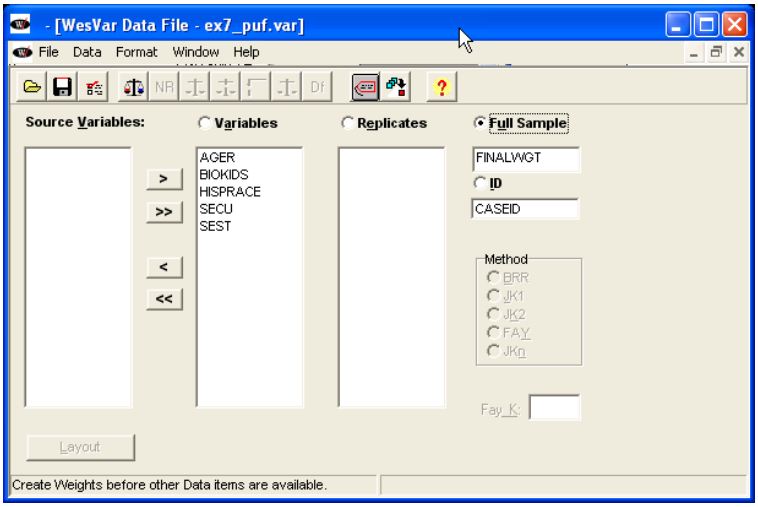
```
. svytab hisprace biokidsx, row se deff percent
pweight:  finalwgt                            Number of obs      =      3807
Strata:   sest                                Number of strata   =        84
PSU:      secu                                Number of PSUs     =       168
                                              Population size    =  50938980

Race and  |
hispanic  |         biokidsx
origin    |  1 or more   No
          |  Children    Children    Total
----------+---------------------------------
 Hispanic |   65.24         34.76      100
          | (2.448)       (2.448)
          |   23.37         41.48
          |
    White |   52.69         47.31      100
          | (2.104)       (2.104)
          |   7.447         7.867
          |
    Black |    60.1          39.9      100
          | (2.294)       (2.294)
          |   31.39         46.15
          |
    Other |    53.4          46.6      100
          | (4.278)       (4.278)
          |   221.7           253
          |
    Total |   55.68         44.32      100
          | (1.728)       (1.728)
          |   4.603         4.603

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

  Pearson:
    Uncorrected   chi2(3)         =   36.7049
    Design-based  F(2.80, 235.16) =    8.6825     P = 0.0000

  Mean generalized deff           =    1.4289
  CV of generalized deffs         =    0.2704
```

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.

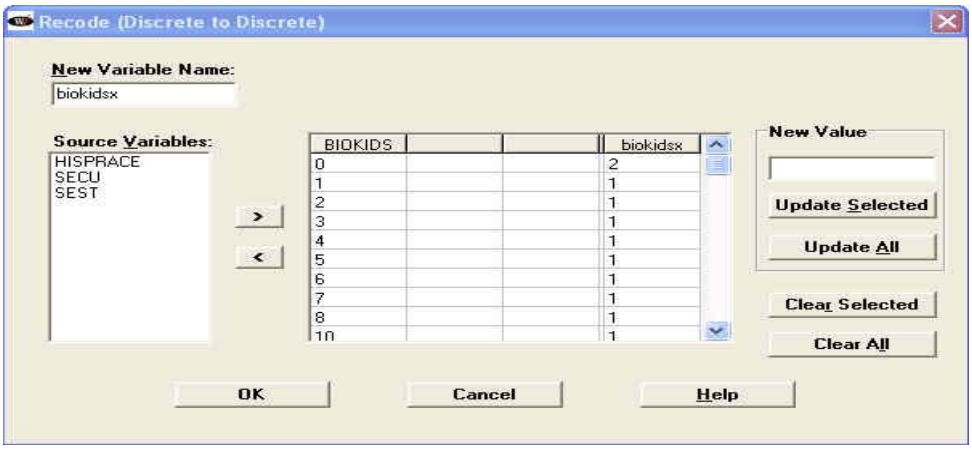
Window 1 displays the selection and categorization of variables to be used in the current analysis. After variables are selected and categorized, a new dataset is created.

WesVar 4.1 Program Window 1



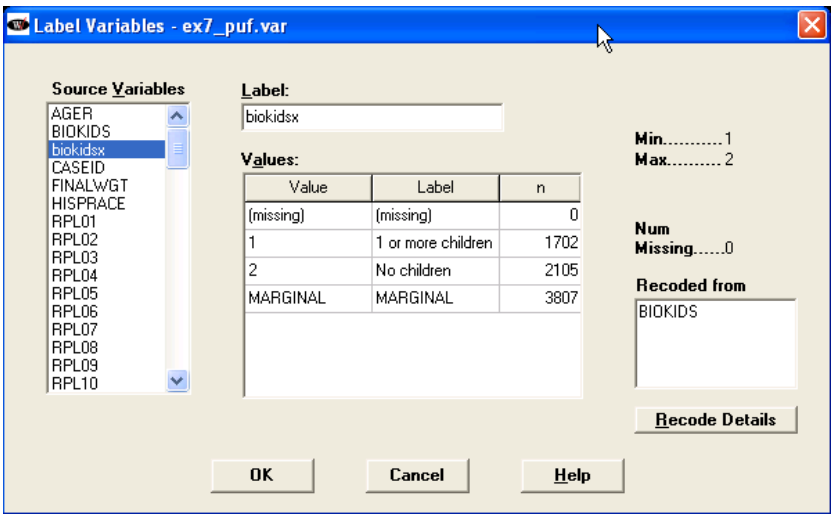
Window 2 displays the procedure for recoding BIOKIDS into 'biokidsx'. To create 'biokidsx', select *Recode* under the *Format* menu and then select the *New Discrete to Discrete* button. After the recodes are created, a new dataset is created.

WesVar 4.1 Program Window 2

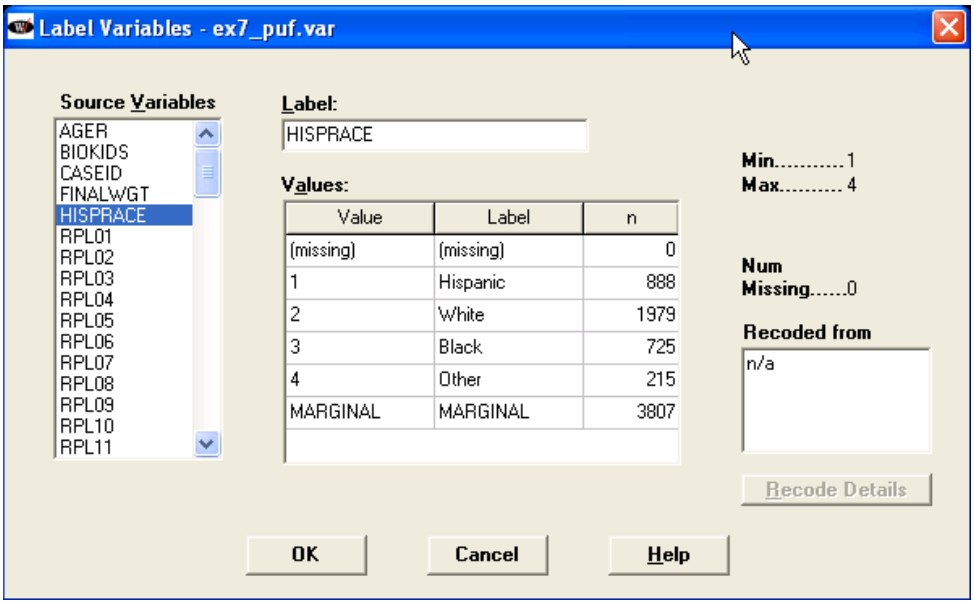


Windows 3a and 3b display how value labels were applied to 'biokidsx' and HISPRACE.

WesVar 4.1 Program Window 3a

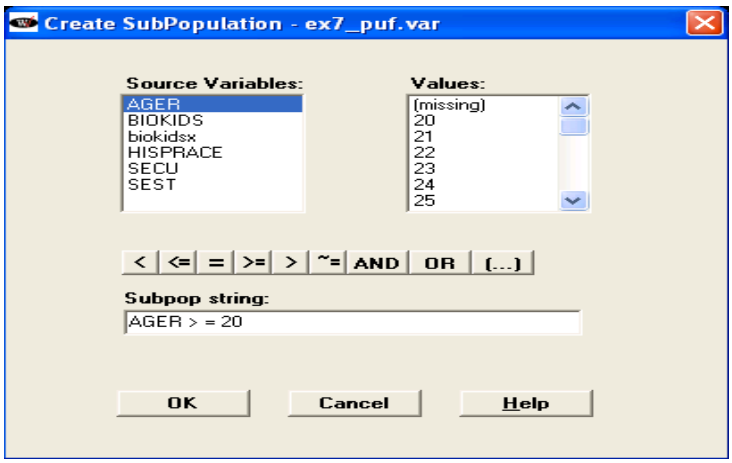


WesVar 4.1 Program Window 3b



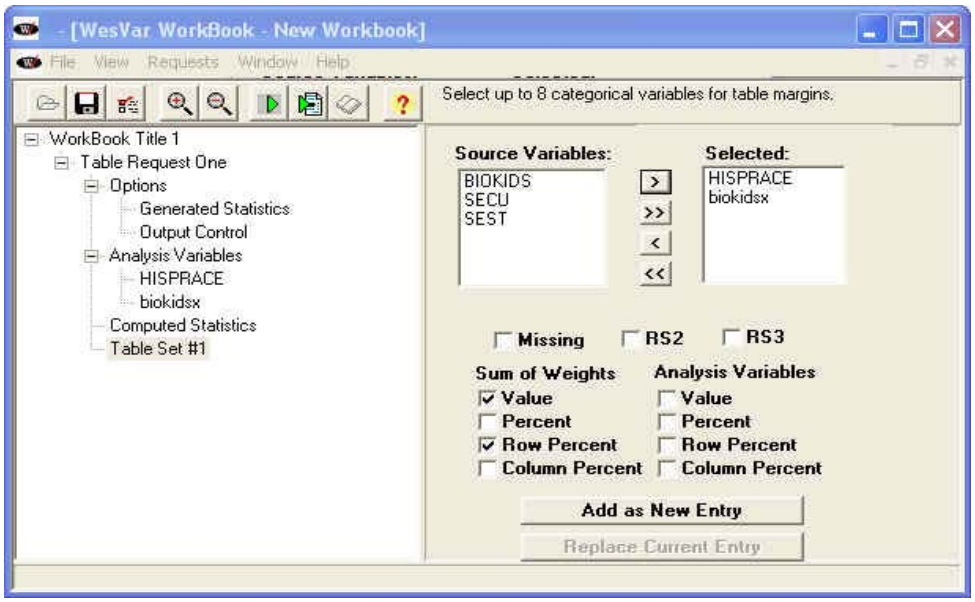
To restrict the analysis to men 20-44 years of age, select *Subset Population* under the *Data* menu.

WesVar 4.1 Program Window 4



From Window 5 select the variables for analysis for a table estimating the percentage of men having one or more children by race and Hispanic origin. Under the *Sum of Weights*, *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. The estimated proportion of males with one or more children by race and Hispanic origin is identical to the other software systems.

WesVar 4.1 Output

WESVAR VERSION NUMBER : v4.1
TIME THE JOB EXECUTED : 12:32:36 10/06/2004
INPUT DATASET NAME : ex7.var
TIME THE INPUT DATASET CREATED : 12:31:28 10/06/2004
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 : HISPRACE, biokidsx
COMPUTED STATISTIC : None Specified.
TABLE(S) : HISPRACE*biokidsx

FACTOR(S) : 1.00

NUMBER OF REPLICATES : 84
NUMBER OF OBSERVATIONS READ : 4928
WEIGHTED NUMBER OF OBSERVATIONS READ : 61147021.513

WesVar 4.1 Output Cont.

HISPRACE	biokidsx	STATISTIC	EST_TYPE	ESTIMATE	STDERROR	CELL_n	DENOM_n	DEFF
Hispanic	1 or more children	SUM_WTS	VALUE	5584669.88	366544.073	509	N/A	N/A
Hispanic	No children	SUM_WTS	VALUE	2975728.24	250412.458	379	N/A	N/A
Hispanic	MARGINAL	SUM_WTS	VALUE	8560398.12	438908.726	888	N/A	N/A
White	1 or more children	SUM_WTS	VALUE	17637646.34	953904.266	737	N/A	N/A
White	No children	SUM_WTS	VALUE	15839852.25	1105233.524	1242	N/A	N/A
White	MARGINAL	SUM_WTS	VALUE	33477498.59	1522376.095	1979	N/A	N/A
Black	1 or more children	SUM_WTS	VALUE	3488507.03	247522.785	378	N/A	N/A
Black	No children	SUM_WTS	VALUE	2315754.17	196635.74	347	N/A	N/A
Black	MARGINAL	SUM_WTS	VALUE	5804261.2	353442.857	725	N/A	N/A
Other	1 or more children	SUM_WTS	VALUE	1653841.32	215889.407	78	N/A	N/A
Other	No children	SUM_WTS	VALUE	1442980.75	207101.935	137	N/A	N/A
Other	MARGINAL	SUM_WTS	VALUE	3096822.07	329851.104	215	N/A	N/A
MARGINAL	1 or more children	SUM_WTS	VALUE	28364664.57	1080895.648	1702	N/A	N/A
MARGINAL	No children	SUM_WTS	VALUE	22574315.42	1362191.692	2105	N/A	N/A
MARGINAL	MARGINAL	SUM_WTS	VALUE	50938979.99	1773994.153	3807	N/A	N/A
Hispanic	1 or more children	SUM_WTS	ROWPCT	65.24	2.48	509	888	2.408
Hispanic	No children	SUM_WTS	ROWPCT	34.76	2.48	379	888	2.408
Hispanic	MARGINAL	SUM_WTS	ROWPCT	100	.	888	888	.
White	1 or more children	SUM_WTS	ROWPCT	52.69	2.11	737	1979	3.535
White	No children	SUM_WTS	ROWPCT	47.31	2.11	1242	1979	3.535
White	MARGINAL	SUM_WTS	ROWPCT	100	.	1979	1979	.
Black	1 or more children	SUM_WTS	ROWPCT	60.1	2.296	378	725	1.594
Black	No children	SUM_WTS	ROWPCT	39.9	2.296	347	725	1.594
Black	MARGINAL	SUM_WTS	ROWPCT	100	.	725	725	.
Other	1 or more children	SUM_WTS	ROWPCT	53.4	4.297	78	215	1.595
Other	No children	SUM_WTS	ROWPCT	46.6	4.297	137	215	1.595
Other	MARGINAL	SUM_WTS	ROWPCT	100	.	215	215	.
MARGINAL	1 or more children	SUM_WTS	ROWPCT	55.68	1.725	1702	3807	4.592
MARGINAL	No children	SUM_WTS	ROWPCT	44.32	1.725	2105	3807	4.592
MARGINAL	MARGINAL	SUM_WTS	ROWPCT	100	.	3807	3807	.