Example 2: Variance Estimates for Means using SAS (9.4) and STATA (14)

Mean Number of Children Ever Born, by Hispanic Origin and Race for Women 15-49 Years of Age

Following are SAS and STATA programs and output for an analysis of the mean number of children
born to women 15-49 years of age in the 2017-2019 NSFG female respondent file, by Hispanic origin and race.

The estimates and standard errors are equivalent across SAS and STATA.

In these programs, library and file names are generic; the user must apply names specific to their computing
environment. Formatting and library options are not presented since preferences will vary across user
organizations.

SAS 9.4

The DATA step creates a dataset for females that contains the variables to be used in the analysis. The PROC
SURVEYMEANS step produces a table of weighted means for the variable specified in the VAR statement
(PARITY) by Hispanic origin and race (HISPRACE2) by using the DOMAIN statement. The WEIGHT statement
identifies the weight variable (WGT2017_2019) to be used in estimating the means. PROC SURVEYMEANS
calculates standard errors appropriate to the complex sample design variables specified in the STRATUM and
CLUSTER statements. The NOMCAR option is included in this PROC SURVEYMEANS example even though
there are no missing values. Data users should consult official SAS documentation for more information about
the NOMCAR option.

SAS Program

```
data NSFG.EX2;
set NSFG.FEMALES (keep=CASEID SEST SECU HISPRACE2 PARITY WGT2017_2019);
run;

proc surveymeans nomcar;
stratum SEST;
cluster SECU;
domain HISPRACE2;
var PARITY;
weight WGT2017_2019;
run;
```

SAS Output

Mean Numbers of Children Ever Born (PARITY) by Hispanic Origin and Race

The SURVEYMEANS Procedure

Data Summary

Number of Strata	18
Number of Clusters	72
Number of Observations	6141
Sum of Weights	72671926

Statistics

Variable	Label	N	Mean	Std Error of Mean	95% CL for Mean
PARITY	CAPI-based total number of live births (accounting for mult birth) (RECODE)	6141	1.217094	0.039673	1.13755441 1.29663324

Statistics for HISPRACE2 Domains

HISPRACE2	Variable	Label	N	Mean	Std Error of Mean	95% CL for Mean
Hispanic	PARITY	CAPI-based total number of live births (accounting for mult birth) (RECODE)	1711	1.544184	0.082587	1.37860745 1.70976115
Non-Hispanic White, Single Race	PARITY	CAPI-based total number of live births (accounting for mult birth) (RECODE)	2704	1.077374	0.037952	1.00128470 1.15346243
Non-Hispanic Black, Single Race	PARITY	CAPI-based total number of live births (accounting for mult birth) (RECODE)	1191	1.400032	0.070466	1.25875678 1.54130748
Non-Hispanic Other or Multiple Race	PARITY	CAPI-based total number of live births (accounting for mult birth) (RECODE)	535	1.067237	0.096495	0.87377726 1.26069755

STATA 14.0

The *use* statement specifies the dataset to be used. The *svyset* command specifies the weight (WGT2017_2019), strata (SEST), and cluster (SECU) variables to be used by STATA in estimation. These settings are saved for the current session but can be cleared by entering the *clear* command.

The *svy: mean* command produces estimated weighted means for each of the levels of the by variable HISPRACE2 to show means separately by Hispanic origin and race by using the over statement. As with most programming, there are multiple options to get the results you need. For example, STATA also has the option to use a *subpop* command within *svy: mean* (svy, subpop(varname): mean varname). The estimates provided are appropriate to the complex sample design identified by the *svyset* command.

STATA Program

```
use "EX2.DTA"

svyset [pweight=WGT2017_2019], strata(SEST) psu(SECU)

svy: mean PARITY, over(HISPRACE2)
```

STATA Output

```
. svy: mean PARITY, over(HISPRACE2)
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =        18        Number of obs   =        6,141
Number of PSUs   =        72        Population size =   72,671,926
                                    Design df       =           54

     Hispanic: HISPRACE2 = Hispanic
    _subpop_2: HISPRACE2 = Non-Hispanic White, Single Race
    _subpop_3: HISPRACE2 = Non-Hispanic Black, Single Race
    _subpop_4: HISPRACE2 = Non-Hispanic Other or Multiple R
```

		Linearized		
Over	Mean	Std. Err.	[95% Conf.	Interval]
PARITY				
Hispanic	1.544184	.0825869	1.378607	1.709761
_subpop_2	1.077374	.0379518	1.001285	1.153462
_subpop_3	1.400032	.0704658	1.258757	1.541307
_subpop_4	1.067237	.0964947	.8737773	1.260698