

Example 1: Variance Estimates for Percentages using SAS (9.4) and STATA (14)

Percentage of Women Ages 15-44 Currently Using the Oral Contraceptive Pill, by Age

Following are SAS and STATA programs and output for an analysis of the percentage of women in the 2013-2015 NSFG who were using the oral contraceptive pill during the month of interview. A cross-tabulation of use of the pill by age (in six categories: 15-19, 20-24, 25-29, 30-34, and 40-44) is generated.

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

In these programs, variables in upper case represent variables as named on the data files. Variables in lower case represent variables that were created as part of this program. Library and file names are generic; the user must apply names specific to his or her computing environment. Formatting and library options have been deleted since preferences will vary across user organizations. SAS format statements could be used instead of creating new variables for some examples shown here.

SAS 9.4

The DATA and SET steps create a dataset for females which contains the variables to be used in the analysis, age categories (agerx) and current use of contraceptive pill (cpill). The PROC SURVEYFREQ produces a cross-tabulation of unweighted and weighted cell counts for the variables (i.e. agerx by cpill) specified in the TABLE statement. The WEIGHT statement identifies the weight variable WGT2013_2015. 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.

SAS Program

```
data EX1;
set NSFG.FEMALES;

if 15 le AGER le 19 then agerx=1;
else if 20 le AGER le 24 then agerx=2;
else if 25 le AGER le 29 then agerx=3;
else if 30 le AGER le 34 then agerx=4;
else if 35 le AGER le 39 then agerx=5;
else if AGER ge 40 then agerx=6;

/*Value of 6 on CONSTAT1 is oral contraceptive pill;
if CONSTAT1=6 then cpill=1;
else cpill=2;
run;

proc surveyfreq;
stratum SEST;
cluster SECU;
weight WGT2013_2015;
table agerx*cpill /ROW NOCELLPERCENT nospase;
format agerx agerf. cpill pillf.;
run;
```

SAS Output

NSFG 2013-2015 Percentage of Women Currently Using the Pill by Age

The SURVEYFREQ Procedure

Data Summary

Number of Strata	18
Number of Clusters	72
Number of Observations	5699
Sum of Weights	61491766

Table of agerx by cpill

agerx	cpill	Frequency	Weighted Frequency	Std Err of Wgt Freq	Row Percent	Std Err of Row Percent

15-19	yes	167	1524464	168986	16.1723	1.5414
	no	843	7901918	473135	83.8277	1.5414
	Total	1010	9426382	536559	100.000	

20-24	yes	216	2913058	379830	27.8173	2.4162
	no	737	7559067	579838	72.1827	2.4162
	Total	953	10472125	815825	100.000	

25-29	yes	180	2007617	218458	18.7197	1.9890
	no	867	8717006	547078	81.2803	1.9890
	Total	1047	10724623	569295	100.000	

30-34	yes	115	1430570	173378	13.4367	1.5074
	no	920	9216141	629796	86.5633	1.5074
	Total	1035	10646711	678580	100.000	

35-39	yes	86	901038	147446	9.0851	1.5222
	no	787	9016730	581215	90.9149	1.5222
	Total	873	9917768	578727	100.000	

40-44	yes	55	795730	197084	7.7224	1.6053
	no	726	9508427	697398	92.2776	1.6053
	Total	781	10304157	803784	100.000	

Total	yes	819	9572477	633971		
	no	4880	51919289	2245834		

Table of agerx by cpill

agerx	cpill	Frequency	Weighted Frequency	Std Err of Wgt Freq	Row Percent	Std Err of Row Percent
Total	Total	5699	61491766	2611759		

STATA 14

The *use* statement specifies the dataset to be used. The *svyset* command specifies the weight (WGT2013_2015), 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 or running *svyset* again with different settings. The *generate* and *replace* statements create the recoded variables *agerx* and *cpill*. The *svytab* command produces a cross-tabulation of *agerx* and *cpill* and provides estimates appropriate to the complex sample design identified by the *svyset* command. The requested estimates and output are limited by specifying *row* and *se* after the *svytab* command.

STATA Program

```
use "EX1.DTA"

svyset [pweight=WGT2013_2015], strata(SEST) psu(SECU)

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 cpill=2
replace cpill=1 if CONSTAT1==6

svy: tab agerx cpill, row se percent
```

STATA Output

```
. svy: tab agerx cpill, row se percent
(running tabulate on estimation sample)
```

```
Number of strata   =      18           Number of obs     =     5,699
Number of PSUs    =      72           Population size   =  61,491,766
Design df         =                   =      54
```

agerx	cpill		Total
	yes	no	
15-19	16.17 (1.541)	83.83 (1.541)	100
20-24	27.82 (2.416)	72.18 (2.416)	100
25-29	18.72 (1.989)	81.28 (1.989)	100
30-34	13.44 (1.507)	86.56 (1.507)	100
35-39	9.085 (1.522)	90.91 (1.522)	100
40-44	7.722 (1.605)	92.28 (1.605)	100
Total	15.57 (.7764)	84.43 (.7764)	100

```
Key: row percentage
(linearized standard error of row percentage)
```

```
Pearson:
Uncorrected chi2(5) = 196.0721
Design-based F(4.45, 240.39) = 15.6285 P = 0.0000
```