

SIPP Complex Sample Specification for SAS, Stata, and WesVar

by Martin C. Seay¹ and Robert B. Nielsen²

SAS:

```
proc survey_____ data=your_datafile;
  *your_secondary_command and your variable(s) or model statement;
  strata      GVARSTR ;          * this is the variance strata variable;
  cluster     GHLFSAM ;         * this is the half sample variable;
  weight      _yourwgt_;       * person/family/household weight variable;
run;
```

Change the **proc survey_____** for other procedures. For example:

```
proc surveylogistic data=your_datafile;          *(Taylor series method);
  model        DEPENDENT (event='1')= INDEPENDENT;
  strata      GVARSTR ;          * this is the variance strata variable;
  cluster     GHLFSAM ;         * this is the half sample variable;
  weight      _yourwgt_;       * person/family/household weight variable;
run;
```

```
proc surveylogistic data=your_datafile varmethod=BRR(fay); *(Replicate weights);
  model        DEPENDENT (event='1')= INDEPENDENT;
  weight      _yourwgt_;       * person/family/household weight variable;
  repweights  REPWT_1-REPWT_N ; * replicate weights 1 to n ;
run;
```

WesVar:

To generate replicate weights (when using 1996 panel or earlier)

1. Open a SAS dataset with WesVar.
2. Specify the regular sample weight in "Full Sample" frame.
3. Bring all other variables into "variables" frame. No ID variable needed.
4. Save the dataset to convert to WesVar formatted dataset.
5. After dataset created, click on "Create Weights" button.
6. For "Method" click "Fay" and use .5 for "Fay K" type.
7. For "VarStrat" use GVARSTR; for "VarUnit" use GHLFSAM.
8. Note that the Hadamard matrix is automatically specified.
9. No other frames need to be filled...click OK to create replicate weights.
10. Wesvar supports few procedures (bivariate estimates plus logistic and OLS regression) but you may export the replicate weights for use elsewhere.
11. Note that SIPP replicate weights can be downloaded directly from the SIPP FTP site for the 2001 panels and later. WesVar can read these pre-calculated replicate weights if you desire to complete your analyses in WesVar. However, we believe it is advantageous to use WesVar only to generate replicate weights then use those weights in SAS or Stata because SAS and Stata offer more analytic tools and procedures.

¹ Assistant Professor, School of Family Studies and Human Services, 318 Justin Hall, Kansas State University, Manhattan, KS. 66506. Email: mseay@ksu.edu

² Associate Professor, Department of Housing and Consumer Economics, University of Georgia, 205 Consumer Research Center, Athens, GA 30602. Email: rnielsen@uga.edu

Stata: (using the estimation of a mean value as our example)

Taylor series linearization method using menus

Sampling unit is GHLFSAM; Strata is GVARSTR.
 Sampling weight variable is your regular full sample pweight.
 Standard error can be specified as linearized (for Taylor) or other.
 Within the procedure, specify "survey data estimation" in SE.

Taylor series linearization method using code

```
First svyset GHLFSAM [pweight=_yourwgt_], strata(GVARSTR) vce(linearized)
Then svy: mean _yourvariable_
```

Replicate weight method using menus

Sampling weight variable is your full sample pweight.
 Balanced repeated replicate(BRR) weight variables are REPWT_1-REPWT_N
 Fay's adjustment is .5
 Within the procedure, specify "survey data estimation" in SE.

Replicate weight method using code

```
First svyset [pweight=_yourwgt_], brrweight(REPWT_1-REPWT_N) fay(.5)
vce(brr)
Then svy: mean _yourvariable_
```

For further information on the importance of incorporating SIPP complex sampling design in analyses see Nielsen et al. (2009); technical documentation can be found on the Census Bureau SIPP page (n.d.).

References

- Nielsen, R. B., Davern, M., Jones, A. Jr., & Boies, J. L. (2009). Complex sample design effects and health insurance variance estimation. *Journal of Consumer Affairs*, 43(2), 346-366.
- U.S. Census Bureau. (n.d). Technical documentation: Survey of Income and Program Participation (SIPP). Washington, DC. Available:
<http://www.census.gov/apds/techdoc/sipp/sipp.html>

Suggested citation:

Seay, M. C., & Nielsen, R. B. (2012). *SIPP complex sample specification for SAS, Stata, and WesVar*. Technical note, Department of Housing and Consumer Economics, University of Georgia, Athens, GA.