svy sdr — Successive difference replication for survey data

Description

`svy sdr` performs successive difference replication (SDR) estimation of specified statistics (or expressions) for a Stata command or a user-written program. The command is executed once for each replicate using sampling weights that are adjusted according to the SDR methodology. Any Stata estimation command listed in [SVY] **svy estimation** may be used with `svy sdr`. User-written programs that meet the requirements in [P] **program properties** may also be used.

Quick start

Estimate population mean of `v1` using SDR standard-error estimates with sampling weight `wvar1` and replicate weights in variables with prefix `rwvar`

```
svyset [pweight=wvar1], sdrweight(rwvar*)
svy sdr _b: mean v1
```

Same as above

```
svyset [pweight=wvar1], sdrweight(rwvar*) vce(sdr)
svy: mean v1
```

SDR estimate of the standard error of the difference between the means of `v2` and `v3` using either `svyset` command above

```
svy sdr (_b[v2]-_b[v3]): mean v2 v3
```

Same as above, but name the result `diff` and save results from each replication to `mydata.dta`

```
svy sdr diff=(_b[v2]-_b[v3]), saving(mydata): mean v2 v3
```

Same as above

```
sdr diff=(_b[v2]-_b[v3]), saving(mydata): mean v2 v3
```

Note: Any estimation command meeting the requirements specified in the *Description* may be substituted for `mean` in the examples above.

Menu

Statistics > Survey data analysis > Resampling > Successive difference replications estimation

Syntax

[svy] sdr *exp_list* [, *svy_options sdr_options eform_option*] : *command*

svy_options	Description
if/in	
subpop([*varname*] [*if*])	identify a subpopulation
Reporting	
level(#)	set confidence level; default is level(95)
noheader	suppress table header
nolegend	suppress table legend
noadjust	do not adjust model Wald statistic
nocnsreport	do not display constraints
display_options	control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
coeflegend	display legend instead of statistics

coeflegend is not shown in the dialog boxes for estimation commands.

sdr_options	Description
Options	
saving(*filename* [, ...])	save results to *filename*; save statistics in double precision; save results to *filename* every # replications
mse	use MSE formula for variance
Reporting	
verbose	display the full table legend
nodots	suppress replication dots
dots(#)	display dots every # replications
noisily	display any output from *command*
trace	trace *command*
title(*text*)	use *text* as title for SDR results
Advanced	
nodrop	do not drop observations
reject(*exp*)	identify invalid results
dof(#)	design degrees of freedom

svy requires that the survey design variables be identified using svyset; see [SVY] **svyset**.

command defines the statistical command to be executed. The by prefix cannot be part of *command*.

collect is allowed; see **[U] 11.1.10 Prefix commands**.

See **[U] 20 Estimation and postestimation commands** for more capabilities of estimation commands.

Warning: Using if or in restrictions will often not produce correct variance estimates for subpopulations. To compute estimates for subpopulations, use the subpop() option.

svy sdr requires that the successive difference replicate weights be identified using svyset.

exp_list specifies the statistics to be collected from the execution of *command*. *exp_list* is required unless *command* has the `svyb` program property, in which case *exp_list* defaults to `_b`; see [P] **program properties**. The expressions in *exp_list* are assumed to conform to the following:

exp_list contains	(*name*: *elist*)
	elist
	eexp
elist contains	*newvarname* = (*exp*)
	(*exp*)
eexp is	*specname*
	[*eqno*]*specname*
specname is	`_b`
	`_b[]`
	`_se`
	`_se[]`
eqno is	`#` #
	name

exp is a standard Stata expression; see **[U] 13 Functions and expressions**.

Distinguish between `[]`, which are to be typed, and [ ], which indicate optional arguments.

Options

svy_options; see [SVY] **svy**.

Options

`saving(`*filename*[, *suboptions*]`)` creates a Stata data file (`.dta` file) consisting of (for each statistic in *exp_list*) a variable containing the replicates.

`double` specifies that the results for each replication be stored as `doubles`, meaning 8-byte reals. By default, they are stored as `floats`, meaning 4-byte reals. This option may be used without the `saving()` option to compute the variance estimates by using double precision.

`every(#)` specifies that results be written to disk every #th replication. `every()` should be specified in conjunction with `saving()` only when *command* takes a long time for each replication. This will allow recovery of partial results should some other software crash your computer. See [P] **postfile**.

`replace` specifies that *filename* be overwritten if it exists. This option does not appear in the dialog box.

`mse` specifies that `svy sdr` compute the variance by using deviations of the replicates from the observed value of the statistics based on the entire dataset. By default, `svy sdr` computes the variance by using deviations of the replicates from their mean.

Reporting

`verbose` requests that the full table legend be displayed.

`nodots` and `dots(#)` specify whether to display replication dots. By default, one dot character is displayed for each successful replication. An "x" is displayed if *command* returns an error, and an "e" is displayed if at least one value in *exp_list* is missing. You can also control whether dots are displayed using `set dots`; see [R] **set**.

`nodots` suppresses display of the replication dots.

`dots(#)` displays dots every # replications. `dots(0)` is a synonym for `nodots`.

`noisily` requests that any output from *command* be displayed. This option implies the `nodots` option.

`trace` causes a trace of the execution of *command* to be displayed. This option implies the `noisily` option.

`title(`*text*`)` specifies a title to be displayed above the table of SDR results; the default title is "SDR results".

eform_option; see [R] ***eform_option***. This option is ignored if *exp_list* is not `_b`.

Advanced

`nodrop` prevents observations outside `e(sample)` and the `if` and `in` qualifiers from being dropped before the data are resampled.

`reject(`*exp*`)` identifies an expression that indicates when results should be rejected. When *exp* is true, the resulting values are reset to missing values.

`dof(#)` specifies the design degrees of freedom, overriding the default calculation, $\text{df} = N_{\text{psu}} - N_{\text{strata}}$.

Remarks and examples

SDR was first introduced by Fay and Train (1995) as a method of variance estimation for annual demographic supplements to the Current Population Survey (CPS). In SDR, the model is fit multiple times, once for each of a set of adjusted sampling weights. The variance is estimated using the resulting replicated point estimates.

▷ Example 1

The US Census Bureau publishes public-use data from several of its surveys. These data can be downloaded from https://factfinder.census.gov. We downloaded the American Community Survey (ACS) Public Use Microdata Sample (PUMS) data collected in 2007. We extracted data for the state of Texas and kept the variables containing age, sex, and sampling weight for each person in the dataset. This sample dataset also contains 80 SDR weight variables.

```
. use https://www.stata-press.com/data/r19/ss07ptx
. svyset
Sampling weights: pwgtp
             VCE: sdr
             MSE: off
     SDR weights: pwgtp1 .. pwgtp80
     Single unit: missing
        Strata 1: <one>
 Sampling unit 1: <observations>
           FPC 1: <zero>
```

This dataset was already `svyset` as

```
. svyset [pw=pwgtp], sdrweight(pwgtp1-pwgtp80) vce(sdr)
```

Here we estimate the average age of the males and of the females for our Texas subpopulation. The standard errors are estimated using SDR.

```
. svy: mean agep, over(sex)
(running mean on estimation sample)
SDR replications (80): .........10.........20.........30.........40.........50..
> .......60.........70.........80 done
Survey: Mean estimation           Number of obs   =    230,817
                                  Population size = 23,904,380
                                  Replications    =         80
```

	SDR			
	Mean	std. err.	[95% conf.	interval]
c.agep@sex				
Male	33.24486	.0470986	33.15255	33.33717
Female	35.23908	.0386393	35.16335	35.31481

◁

❑ Technical note

When the `svy sdr` prefix is used with a user-defined program and when the expression list is `_b`, `svy sdr` calls

```
set coeftabresults off
```

before entering the replication loop to prevent Stata from performing unnecessary calculations. This means that, provided option `noisily` is not specified, estimation commands will not build or post the coefficient table matrix `r(table)`.

If your program calls an estimation command and needs `r(table)` to exist to perform properly, then your program will need to call

```
set coeftabresults on
```

before calling other estimation commands.

❑

Stored results

In addition to the results documented in [SVY] **svy**, `svy sdr` stores the following in `e()`:

Scalars

`e(N_reps)`	number of replications
`e(N_misreps)`	number of replications with missing values
`e(k_exp)`	number of standard expressions
`e(k_eexp)`	number of _b/_se expressions
`e(k_extra)`	number of extra estimates added to _b

Macros

`e(cmdname)`	command name from *command*
`e(cmd)`	same as `e(cmdname)` or `sdr`
`e(vce)`	`sdr`
`e(exp#)`	#th expression
`e(sdrweight)`	`sdrweight()` variable list

Matrices

`e(b_sdr)`	SDR means
`e(V)`	SDR variance estimates

When *exp_list* is _b, `svy sdr` will also carry forward most of the results already in `e()` from *command*.

Methods and formulas

See [SVY] **Variance estimation** for details regarding SDR variance estimation.

Reference

Fay, R. E., and G. F. Train. 1995. "Aspects of survey and model-based postcensal estimation of income and poverty characteristics for states and counties". In *Proceedings of the Government Statistics Section*, 154–159. American Statistical Association.

Also see

[SVY] **svy postestimation** — Postestimation tools for svy

[SVY] **svy bootstrap** — Bootstrap for survey data

[SVY] **svy brr** — Balanced repeated replication for survey data

[SVY] **svy jackknife** — Jackknife estimation for survey data

[SVY] **Calibration** — Calibration for survey data

[SVY] **Poststratification** — Poststratification for survey data

[SVY] **Subpopulation estimation** — Subpopulation estimation for survey data

[SVY] **Variance estimation** — Variance estimation for survey data

[U] 20 Estimation and postestimation commands