svy sdr — Successive difference replication for survey data

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 column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
```

coefflegend  display legend instead of statistics

coefflegend is not shown in the dialog boxes for estimation commands.
svy sdr — Successive difference replication for survey data

\[svr sdr\] — Successive difference replication for survey data

\[save\] options

\begin{center}
\begin{tabular}{ll}
\textbf{saving} & \texttt{filename[, ...]} \hfill \text{save results to \textit{filename}; save statistics in double precision; save results to \textit{filename} every \# replications} \\
\textbf{mse} & \text{use MSE formula for variance} \\
\end{tabular}
\end{center}

\textbf{Reporting}

\begin{center}
\begin{tabular}{ll}
\textbf{verbose} & \text{display the full table legend} \\
\textbf{nodots} & \text{suppress replication dots} \\
\textbf{noisily} & \text{display any output from \textit{command}} \\
\textbf{trace} & \text{trace \textit{command}} \\
\textbf{title(text)} & \text{use \textit{text} as title for SDR results} \\
\end{tabular}
\end{center}

\textbf{Advanced}

\begin{center}
\begin{tabular}{ll}
\textbf{nodrop} & \text{do not drop observations} \\
\textbf{reject(exp)} & \text{identify invalid results} \\
\textbf{dof(\#)} & \text{design degrees of freedom} \\
\end{tabular}
\end{center}

\textit{svy} requires that the survey design variables be identified using \texttt{svyset}; see [SVY] \texttt{svyset}.

See [\texttt{U} 20 Estimation and postestimation commands] for more capabilities of estimation commands.

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

\textit{svy sdr} requires that the successive difference replicate weights be identified using \texttt{svyset}.

\[exp\_list\] contains \hspace{1cm} \texttt{\texttt{(name: elist)}}

\begin{center}
\begin{tabular}{ll}
\texttt{elist} & \texttt{elist} \\
\texttt{eexp} & \texttt{eexp} \\
\end{tabular}
\end{center}

\[elast\] contains \hspace{1cm} \texttt{newvarname = (exp)}

\begin{center}
\begin{tabular}{ll}
\texttt{(exp)} & \texttt{(exp)} \\
\end{tabular}
\end{center}

\[eexp\] is \hspace{1cm} \texttt{specname}

\begin{center}
\begin{tabular}{ll}
\texttt{[eqno]specname} & \texttt{[eqno]specname} \\
\end{tabular}
\end{center}

\textit{specname} is \hspace{1cm} \_b

\begin{center}
\begin{tabular}{ll}
\texttt{\_b[]} & \texttt{\_b[]} \\
\texttt{\_se} & \texttt{\_se[]} \\
\end{tabular}
\end{center}

\textit{eqno} is \hspace{1cm} \# \\

\[name\] is a standard Stata expression; see [\texttt{U} 13 Functions and expressions].

Distinguish between \([\)], which are to be typed, and \([[]]\), which indicate optional arguments.

\textbf{Menu}

Statistics > Survey data analysis > Resampling > Successive difference replications estimation
svy sdr — Successive difference replication for survey data 3

Description

svy sdr performs successive difference replication (SDR) for complex survey data. Typing

```
.svy sdr exp_list: command
```

executes command once for each replicate, using sampling weights that are adjusted according to the SDR methodology.

command defines the statistical command to be executed. Most Stata commands and user-written programs can be used with svy sdr as long as they follow standard Stata syntax, allow the if qualifier, and allow pweights and iweights; see [U] 11 Language syntax. The by prefix may not be part of command.

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.

Options

svy_options; see [SVY] svy.

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 indicates that filename be overwritten if it exists. This option is not shown on 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 suppresses display of the replication dots. By default, one dot character is printed for each successful replication. A red ‘x’ is printed if command returns with an error, and ‘e’ is printed if one of the values in exp_list is missing.

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”.


**svy sdr** — Successive difference replication for survey data

`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, $df = N_{psu} - N_{strata}$.

---

**Remarks and examples**

`sd` was first introduced by [Fay and Train (1995)](http://example.com) 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 U.S. Census Bureau publishes public-use data from several of its surveys. This data can be downloaded from [http://factfinder.census.gov](http://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 http://www.stata-press.com/data/r13/ss07ptx
. svyset
    pweight: pwgtp
    VCE: sdr
    MSE: off
    sdrweight: pwgtp1 pwgtp2 pwgtp3 pwgtp4 pwgtp5 pwgtp6 pwgtp7 pwgtp8 pwgtp9
    pwgtp10 pwgtp11 pwgtp12 pwgtp13 pwgtp14 pwgtp15 pwgtp16
    (output omitted) pwgtp73 pwgtp74 pwgtp75 pwgtp76 pwgtp77 pwgtp78 pwgtp79
    pwgtp80
    Single unit: missing
    Strata 1: <one>
    SU 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 sdr — Successive difference replication for survey data 5

. svy: mean agep, over(sex)
  (running mean on estimation sample)

SDR replications (80)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>50</th>
</tr>
</thead>
</table>

Survey: Mean estimation
Number of obs = 230817
Population size = 23904380
Replications = 80

Male: sex = Male
Female: sex = Female

<table>
<thead>
<tr>
<th></th>
<th>SDR</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Err.</td>
<td>[95% Conf. Interval]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>agep</td>
<td>Male</td>
<td>33.24486</td>
<td>.0470986</td>
<td>33.15255</td>
<td>33.33717</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>35.23908</td>
<td>.0386393</td>
<td>35.16335</td>
<td>35.31481</td>
<td></td>
</tr>
</tbody>
</table>

 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

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

[U] 20 Estimation and postestimation commands

[SVY] _poststratification_ — Poststratification for survey data

[SVY] _subpopulation estimation_ — Subpopulation estimation for survey data

[SVY] _variance estimation_ — Variance estimation for survey data