svy: tabulate oneway — One-way tables for survey data

Description


Quick start

One-way table showing weighted proportions for categories of v1 using svyset data

svy: tabulate v1

Add 95% confidence intervals and weighted counts

svy: tabulate v1, ci count

As above, and display large counts in a more readable format

svy: tabulate v1 ci count format(%11.3g)

Unweighted numbers of observations and weighted proportions for categories of v2

svy: tabulate v2, obs

Weighted proportions and CIs for categories of v3 in the subpopulation defined by v4 > 40

svy, subpop(if v4>40): tabulate v3, ci

Menu

Statistics > Survey data analysis > Tables > One-way tables
## Syntax

### Basic syntax

```
svy: tabulate varname
```

### Full syntax

```
svy [vcetype] [ , svy_options ] : tabulate varname [ if ] [ in ]
[ , tabulate_options display_items display_options ]
```

### Syntax to report results

```
svy [ , display_items display_options ]
```

### vcetype

<table>
<thead>
<tr>
<th>vcetype</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SE</td>
<td>Taylor-linearized variance estimation</td>
</tr>
<tr>
<td>linearized</td>
<td>bootstrap variance estimation; see [SVY] svy bootstrap</td>
</tr>
<tr>
<td>bootstrap</td>
<td>BRR variance estimation; see [SVY] svy brr</td>
</tr>
<tr>
<td>brr</td>
<td>jackknife variance estimation; see [SVY] svy jackknife</td>
</tr>
<tr>
<td>jackknife</td>
<td>SDR variance estimation; see [SVY] svy sdr</td>
</tr>
<tr>
<td>sdr</td>
<td></td>
</tr>
</tbody>
</table>

Specifying a vcetype overrides the default from svyset.

### svy_options

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

### if/in

```
subpop( [varname] [ if ] )
```

identify a subpopulation

### SE

```
bootstrap_options
brr_options
jackknife_options
sdr_options
```

more options allowed with bootstrap variance estimation; see [SVY] bootstrap_options
more options allowed with BRR variance estimation; see [SVY] brr_options
more options allowed with jackknife variance estimation; see [SVY] jackknife_options
more options allowed with SDR variance estimation; see [SVY] sdr_options

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

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.
### tabulate_options

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>stdize(varname)</strong></td>
</tr>
<tr>
<td><strong>stdweight(varname)</strong></td>
</tr>
<tr>
<td><strong>tab(varname)</strong></td>
</tr>
<tr>
<td><strong>missing</strong></td>
</tr>
</tbody>
</table>

### display_items

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>cell</strong></td>
</tr>
<tr>
<td><strong>count</strong></td>
</tr>
<tr>
<td><strong>se</strong></td>
</tr>
<tr>
<td><strong>ci</strong></td>
</tr>
<tr>
<td><strong>deff</strong></td>
</tr>
<tr>
<td><strong>deft</strong></td>
</tr>
<tr>
<td><strong>cv</strong></td>
</tr>
<tr>
<td><strong>srssubpop</strong></td>
</tr>
<tr>
<td><strong>obs</strong></td>
</tr>
</tbody>
</table>

When any of `se`, `ci`, `deff`, `deft`, `cv`, or `srssubpop` is specified, only one of `cell` or `count` can be specified. If none of `se`, `ci`, `deff`, `deft`, `cv`, or `srssubpop` is specified, both `cell` and `count` can be specified.

### display_options

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>level(#)</strong></td>
</tr>
<tr>
<td><strong>prop</strong></td>
</tr>
<tr>
<td><strong>percent</strong></td>
</tr>
<tr>
<td><strong>nomarginal</strong></td>
</tr>
<tr>
<td><strong>nolabel</strong></td>
</tr>
<tr>
<td><strong>cellwidth(#)</strong></td>
</tr>
<tr>
<td><strong>csepwidth(#)</strong></td>
</tr>
<tr>
<td><strong>stubwidth(#)</strong></td>
</tr>
<tr>
<td><strong>format(%,fmt)</strong></td>
</tr>
</tbody>
</table>

Proportion is not shown in the dialog box.

### Options

- **svy_options**: see `[SVY] svy`.

- **stdize(varname)** specifies that the point estimates be adjusted by direct standardization across the strata identified by `varname`. This option requires the `stdweight()` option.

- **stdweight(varname)** specifies the weight variable associated with the standard strata identified in the `stdize()` option. The standardization weights must be constant within the standard strata.
tab(varname) specifies that counts be cell totals of this variable and that proportions (or percentages) be relative to (that is, weighted by) this variable. For example, if this variable denotes income, then the cell “counts” are instead totals of income for each cell, and the cell proportions are proportions of income for each cell.

missing specifies that missing values of varname be treated as another row category rather than be omitted from the analysis (the default).

cell requests that cell proportions (or percentages) be displayed. This is the default if count is not specified.

count requests that weighted cell counts be displayed.

se requests that the standard errors of cell proportions (the default) or weighted counts be displayed. When se (or ci, deff, deft, or cv) is specified, only one of cell or count can be selected. The standard error computed is the standard error of the one selected.

ci requests confidence intervals for cell proportions or weighted counts.

deff and deft request that the design-effect measures DEFF and DEFT be displayed for each cell proportion or weighted count. See [SVY] estat for details.

The deff and deft options are not allowed with estimation results that used direct standardization or poststratification.

cv requests that the coefficient of variation be displayed for each cell proportion, count, or row or column proportion. See [SVY] estat for details.

srssubpop requests that DEFF and DEFT be computed using an estimate of SRS (simple random sampling) variance for sampling within a subpopulation. By default, DEFF and DEFT are computed using an estimate of the SRS variance for sampling from the entire population. Typically, srssubpop would be given when computing subpopulation estimates by strata or by groups of strata.

obs requests that the number of observations for each cell be displayed.

level(#) specifies the confidence level, as a percentage, for confidence intervals. The default is level(95) or as set by set level; see [U] 20.8 Specifying the width of confidence intervals.

percent requests that percentages be displayed instead of proportions.

nomarginal requests that the column marginal not be displayed.

noformat requests that variable labels and value labels be ignored.

cellwidth(#), csepwidth(#), and stubwidth(#) specify widths of table elements in the output; see [P] tabdisp. Acceptable values for the stubwidth() option range from 4 to 32.

format(%fmt) specifies a format for the items in the table. The default is format(%6.0g). See [U] 12.5 Formats: Controlling how data are displayed.

svy: tabulate uses the tabdisp command (see [P] tabdisp) to produce the table. Only five items can be displayed in the table at one time. The ci option implies two items. If too many items are selected, a warning will appear immediately. To view more items, redisplay the table while specifying different options.
Remarks and examples

Despite the long list of options for \texttt{svy: tabulate}, it is a simple command to use. Using the \texttt{svy: tabulate} command is just like using \texttt{tabulate} to produce one-way tables for ordinary data. The main difference is that \texttt{svy: tabulate} computes standard errors appropriate for complex survey data.

Standard errors and confidence intervals can optionally be displayed for weighted counts or cell proportions. The confidence intervals for proportions are constructed using a logit transform so that their endpoints always lie between 0 and 1; see [SVY] \texttt{svy: tabulate twoway}. Associated design effects (DEFF and DEFT) can be viewed for the variance estimates.

\section*{Example 1}

Here we use \texttt{svy: tabulate} to estimate the distribution of the \texttt{race} category variable from our NHANES II dataset (McDowell et al. 1981). Before calling \texttt{svy: tabulate}, we use \texttt{svyset} to declare the survey structure of the data.

\begin{verbatim}
use https://www.stata-press.com/data/r16/nhanes2b
.svysset psuid [pweight=finalwgt], strata(stratid)
pweight: finalwgt VCE: linearized
Single unit: missing Strata 1: stratid
SU 1: psuid FPC 1: <zero>
.svy: tabulate race
(running tabulate on estimation sample)

Number of strata = 31 Number of obs = 10,351
Number of PSUs = 62 Population size = 117,157,513
Design df = 31

1=white, 2=black, 3=other proportion

<table>
<thead>
<tr>
<th></th>
<th>proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>.8792</td>
</tr>
<tr>
<td>Black</td>
<td>.0955</td>
</tr>
<tr>
<td>Other</td>
<td>.0253</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
</tr>
</tbody>
</table>

Key: proportion = cell proportion
\end{verbatim}

Here we display weighted counts for each category of \texttt{race} along with the 95\% confidence bounds, as well as the design effects DEFF and DEFT. We also use the \texttt{format()} option to improve the look of the table.
. svy: tabulate race, format(%11.3g) count ci deff deft  
(running `tabulate` on estimation sample)

<table>
<thead>
<tr>
<th></th>
<th>count</th>
<th>lb</th>
<th>ub</th>
<th>deff</th>
<th>deft</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>102999549</td>
<td>97060400</td>
<td>108938698</td>
<td>60.2</td>
<td>7.76</td>
</tr>
<tr>
<td>Black</td>
<td>11189236</td>
<td>8213964</td>
<td>14164508</td>
<td>18.6</td>
<td>4.31</td>
</tr>
<tr>
<td>Other</td>
<td>2968728</td>
<td>414930</td>
<td>5522526</td>
<td>47.9</td>
<td>6.92</td>
</tr>
<tr>
<td>Total</td>
<td>117157513</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Key:  
- `count` = weighted count  
- `lb` = lower 95% confidence bound for weighted count  
- `ub` = upper 95% confidence bound for weighted count  
- `deff` = deff for variance of weighted count  
- `deft` = deft for variance of weighted count

From the above results, we can conclude with 95% confidence that the number of people in the population that fall within the `White` category is between 97,060,400 and 108,938,698.

### Stored results

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

**Scalars**
- `e(r)` number of rows  
- `e(total)` weighted sum of `tab()` variable

**Macros**
- `e(cmd)` `tabulate`  
- `e(tab)` `tab()` variable  
- `e(rowlab)` label or empty  
- `e(rowvlab)` row variable label  
- `e(rowvar)` `varname`, the row variable  
- `e(setype)` cell or count

**Matrices**
- `e(Prop)` matrix of cell proportion  
- `e(Obs)` matrix of observation count  
- `e(Deff)` DEFF vector for `e(setype)` items  
- `e(Deft)` DEFT vector for `e(setype)` items  
- `e(Row)` values for row variable  
- `e(V_row)` variance for row totals  
- `e(V_srs_row)` `V_{srs}` for row totals  
- `e(Deff_row)` DEFF for row totals  
- `e(Deft_row)` DEFT for row totals

### Methods and formulas

See Methods and formulas in `[SVY] svy: tabulate twoway` for a discussion of how table items and confidence intervals are computed. A one-way table is really just a two-way table that has one row or column.
Margaret E. Martin (1912–2012) is best known for her work developing the U.S. Current Population Survey (CPS). Martin was born in New York City and had an early love for mathematics. She received a bachelor’s degree in economics from Barnard College and went on to earn an MA and a PhD in economics from Columbia University. Martin began her career in the midst of the Great Depression, working for a New Deal agency in New York to classify employers covered by the unemployment insurance system. Despite having the third highest score on the qualifying civil service exam, she almost did not take the job because she “had been trained by economists primarily, and they had a very low opinion of government work”.

Her work in New York allowed her to later move to the U.S. Bureau of Budget (now the Office of Management and Budget), where she joined the team that developed the CPS. The majority of Martin’s work focused on the CPS, a survey of employment and demographics among U.S. households. She worked to explain differences in previous unemployment survey results derived from sampling businesses. She also oversaw an effort to improve the reliability of information from the CPS by adding questions that addressed labor-force participation and the use of paid and unpaid leave. Today, the CPS is a continuous monthly survey and the primary source of information about characteristics of the U.S. labor force.

In 1973, Martin became the first executive director of the National Academy of Sciences’ Committee on National Statistics. She was elected president of the American Statistical Association (ASA) in 1980 and was the first recipient of the ASA’s Founders Award.

Reference


Also see

[SVY] `svy postestimation` — Postestimation tools for `svy`
[SVY] `svy` — The survey prefix command
[SVY] `svy: tabulate twoway` — Two-way tables for survey data
[SVY] `svydescribe` — Describe survey data
[SVY] `Calibration` — Calibration for survey data
[SVY] `Direct standardization` — Direct standardization of means, proportions, and ratios
[SVY] `Poststratification` — Poststratification for survey data
[SVY] `Subpopulation estimation` — Subpopulation estimation for survey data
[SVY] `Variance estimation` — Variance estimation for survey data
[R] `tabulate oneway` — One-way table of frequencies
[U] 20 Estimation and postestimation commands