Title

stata.com

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

Description

`svy: tabulate` produces one-way tabulations for complex survey data. See [SVY] **svy: tabulate twoway** for two-way tabulations for complex survey data.

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	Description
SE	
`linearized`	Taylor-linearized variance estimation
`bootstrap`	bootstrap variance estimation; see [SVY] **svy bootstrap**
`brr`	BRR variance estimation; see [SVY] **svy brr**
`jackknife`	jackknife variance estimation; see [SVY] **svy jackknife**
`sdr`	SDR variance estimation; see [SVY] **svy sdr**

Specifying a *vcetype* overrides the default from `svyset`.

svy_options	Description
if/in	
`subpop(`[*varname*] [*if*]`)`	identify a subpopulation
SE	
bootstrap_options	more options allowed with bootstrap variance estimation; see [SVY] ***bootstrap_options***
brr_options	more options allowed with BRR variance estimation; see [SVY] ***brr_options***
jackknife_options	more options allowed with jackknife variance estimation; see [SVY] ***jackknife_options***
sdr_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	Description
Model	
`stdize(`*varname*`)`	variable identifying strata for standardization
`stdweight(`*varname*`)`	weight variable for standardization
`tab(`*varname*`)`	variable for which to compute cell totals/proportions
`missing`	treat missing values like other values

display_items	Description
Table items	
`cell`	cell proportions
`count`	weighted cell counts
`se`	standard errors
`ci`	confidence intervals
`deff`	display the DEFF design effects
`deft`	display the DEFT design effects
`cv`	display the coefficient of variation
`srssubpop`	report design effects assuming SRS within subpopulation
`obs`	cell observations

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	Description
Reporting	
`level(#)`	set confidence level; default is `level(95)`
`proportion`	display proportions; the default
`percent`	display percentages instead of proportions
`nomarginal`	suppress column marginal
`nolabel`	suppress displaying value labels
`cellwidth(#)`	cell width
`csepwidth(#)`	column-separation width
`stubwidth(#)`	stub width
`format(%`*fmt*`)`	cell format; default is `format(%6.0g)`

`proportion` is not shown in the dialog box.

Options

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

Model

`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).

Table items

`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.

Reporting

`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**.

`proportion`, the default, requests that proportions be displayed.

`percent` requests that percentages be displayed instead of proportions.

`nomarginal` requests that the column marginal not be displayed.

`nolabel` 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

stata.com

Despite the long list of options for svy: tabulate, it is a simple command to use. Using the svy: tabulate command is just like using tabulate to produce one-way tables for ordinary data. The main difference is that 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] **svy: tabulate twoway**. Associated design effects (DEFF and DEFT) can be viewed for the variance estimates.

▷ Example 1

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

```
. use http://www.stata-press.com/data/r15/nhanes2b
. svyset 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
-----------+------------
     White |      .8792
     Black |      .0955
     Other |      .0253
           |
     Total |           1
------------------------
  Key:  proportion  =  cell proportion
```

Here we display weighted counts for each category of race along with the 95% confidence bounds, as well as the design effects DEFF and DEFT. We also use the format() option to improve the look of the table.

```
. svy: tabulate race, format(%11.3g) count ci deff deft
(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	count	lb	ub	deff	deft
White	102999549	97060400	108938698	60.2	7.76
Black	11189236	8213964	14164508	18.6	4.31
Other	2968728	414930	5522526	47.9	6.92
Total	117157513				

```
  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(Deff_row)`	DEFF for row totals
`e(V_row)`	variance for row totals
`e(V_srs_row)`	V_{srs} 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

McDowell, A., A. Engel, J. T. Massey, and K. Maurer. 1981. Plan and operation of the Second National Health and Nutrition Examination Survey, 1976–1980. *Vital and Health Statistics* 1(15): 1–144.

Also see

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

[SVY] **svydescribe** — Describe survey data

[R] **tabulate oneway** — One-way table of frequencies

[SVY] **svy: tabulate twoway** — Two-way tables for survey data

[U] **20 Estimation and postestimation commands**

[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] **svy** — The survey prefix command

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