Title

stata.com

total — Estimate totals

Description

`total` produces estimates of totals, along with standard errors.

Quick start

Total of continuous variable `v1`

```
total v1
```

As above, but restrict estimation to observations where `catvar` = 1

```
total v1 if catvar==1
```

As above, but using `svyset` data

```
svy, subpop(if catvar==1): total v1
```

Total of `v1` for each level of `catvar`

```
total v1, over(catvar)
```

With jackknife standard errors

```
total v1, vce(jackknife)
```

Menu

Statistics > Summaries, tables, and tests > Summary and descriptive statistics > Totals

Syntax

```
total varlist [if] [in] [weight] [, options]
```

options	Description
if/in/over	
over(*varlist*[, nolabel])	group over subpopulations defined by *varlist*; optionally, suppress group labels
SE/Cluster	
vce(*vcetype*)	*vcetype* may be analytic, cluster *clustvar*, bootstrap, or jackknife
Reporting	
level(#)	set confidence level; default is level(95)
noheader	suppress table header
nolegend	suppress table legend
display_options	control column formats and line width
coeflegend	display legend instead of statistics

bootstrap, jackknife, mi estimate, rolling, statsby, and svy are allowed; see **[U] 11.1.10 Prefix commands**.
vce(bootstrap) and vce(jackknife) are not allowed with the mi estimate prefix.
Weights are not allowed with the bootstrap prefix; see [R] **bootstrap**.
vce() and weights are not allowed with the svy prefix; see [SVY] **svy**.
fweights, iweights, and pweights are allowed; see **[U] 11.1.6 weight**.
coeflegend does not appear in the dialog box.
See **[U] 20 Estimation and postestimation commands** for more capabilities of estimation commands.

Options

if/in/over

over(*varlist* [, nolabel]) specifies that estimates be computed for multiple subpopulations, which are identified by the different values of the variables in *varlist*.

When this option is supplied with one variable name, such as over(*varname*), the value labels of *varname* are used to identify the subpopulations. If *varname* does not have labeled values (or there are unlabeled values), the values themselves are used, provided that they are nonnegative integers. Noninteger values, negative values, and labels that are not valid Stata names are substituted with a default identifier.

When over() is supplied with multiple variable names, each subpopulation is assigned a unique default identifier.

nolabel specifies that value labels attached to the variables identifying the subpopulations be ignored.

SE/Cluster

vce(*vcetype*) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (analytic), that allow for intragroup correlation (cluster *clustvar*), and that use bootstrap or jackknife methods (bootstrap, jackknife); see [R] ***vce_option***.

`vce(analytic)`, the default, uses the analytically derived variance estimator associated with the sample total.

Reporting

`level(#)`; see [R] **estimation options**.

`noheader` prevents the table header from being displayed. This option implies `nolegend`.

`nolegend` prevents the table legend identifying the subpopulations from being displayed.

display_options: `cformat(%`*fmt*`)` and `nolstretch`; see [R] **estimation options**.

The following option is available with `total` but is not shown in the dialog box:

`coeflegend`; see [R] **estimation options**.

Remarks and examples

stata.com

▷ Example 1

Suppose that we collected data on incidence of heart attacks. The variable `heartatk` indicates whether a person ever had a heart attack (1 means yes; 0 means no). We can then estimate the total number of persons who have had heart attacks for each `sex` in the population represented by the data we collected.

```
. use http://www.stata-press.com/data/r14/total
. total heartatk [pw=swgt], over(sex)
Total estimation                  Number of obs   =      4,946
         Male: sex = Male
       Female: sex = Female
------------------------------------------------------------------------------
             Over |      Total   Std. Err.     [95% Conf. Interval]
------------------+-----------------------------------------------------------
heartatk         |
            Male |     944559   104372.3       739943     1149175
          Female |     581590   82855.59      419156.3    744023.7
------------------------------------------------------------------------------
```

◁

Stored results

total stores the following in e():

Scalars

e(N)	number of observations
e(N_over)	number of subpopulations
e(N_clust)	number of clusters
e(k_eq)	number of equations in e(b)
e(df_r)	sample degrees of freedom
e(rank)	rank of e(V)

Macros

e(cmd)	total
e(cmdline)	command as typed
e(varlist)	*varlist*
e(wtype)	weight type
e(wexp)	weight expression
e(title)	title in estimation output
e(cluster)	name of cluster variable
e(over)	*varlist* from over()
e(over_labels)	labels from over() variables
e(over_namelist)	names from e(over_labels)
e(vce)	*vcetype* specified in vce()
e(vcetype)	title used to label Std. Err.
e(properties)	b V
e(estat_cmd)	program used to implement estat
e(marginsnotok)	predictions disallowed by margins

Matrices

e(b)	vector of total estimates
e(V)	(co)variance estimates
e(_N)	vector of numbers of nonmissing observations
e(error)	error code corresponding to e(b)

Functions

e(sample)	marks estimation sample

Methods and formulas

Methods and formulas are presented under the following headings:

The total estimator

Let y denote the variable on which to calculate the total and $y_j, j = 1, \ldots, n$, denote an individual observation on y. Let w_j be the frequency weight (or iweight or pweight), and if no weight is specified, define $w_j = 1$ for all j. The sum of the weights is an estimate of the population size:

$$\widehat{N} = \sum_{j=1}^{n} w_j$$

If the population values of y are denoted by $Y_j, j = 1, \ldots, N$, the associated population total is

$$Y = \sum_{j=1}^{N} Y_j = N\overline{y}$$

where $\overline{y}$ is the population mean. The total is estimated as

$$\widehat{Y} = \widehat{N}\overline{y}$$

The variance estimator for the total is

$$\widehat{V}(\widehat{Y}) = \widehat{N}^2\widehat{V}(\overline{y})$$

where $\widehat{V}(\overline{y})$ is the variance estimator for the mean; see [R] **mean**. The standard error of the total is the square root of the variance.

If x, x_j, $\overline{x}$, and $\widehat{X}$ are similarly defined for another variable (observed jointly with y), the covariance estimator between $\widehat{X}$ and $\widehat{Y}$ is

$$\widehat{\mathrm{Cov}}(\widehat{X}, \widehat{Y}) = \widehat{N}^2\widehat{\mathrm{Cov}}(\overline{x}, \overline{y})$$

where $\widehat{\mathrm{Cov}}(\overline{x}, \overline{y})$ is the covariance estimator between two means; see [R] **mean**.

Survey data

See [SVY] **variance estimation** and [SVY] **poststratification** for discussions that provide background information for the following formulas.

The survey total estimator

Let Y_j be a survey item for the jth individual in the population, where $j = 1, \ldots, M$ and M is the size of the population. The associated population total for the item of interest is

$$Y = \sum_{j=1}^{M} Y_j$$

Let y_j be the survey item for the jth sampled individual from the population, where $j = 1, \ldots, m$ and m is the number of observations in the sample.

The estimator $\widehat{Y}$ for the population total Y is

$$\widehat{Y} = \sum_{j=1}^{m} w_j y_j$$

where w_j is a sampling weight. The estimator for the number of individuals in the population is

$$\widehat{M} = \sum_{j=1}^{m} w_j$$

The score variable for the total estimator is the variable itself,

$$z_j(\widehat{Y}) = y_j$$

The poststratified total estimator

Let P_k denote the set of sampled observations that belong to poststratum k, and define $I_{P_k}(j)$ to indicate if the jth observation is a member of poststratum k, where $k = 1, \ldots, L_P$ and L_P is the number of poststrata. Also, let M_k denote the population size for poststratum k. P_k and M_k are identified by specifying the `poststrata()` and `postweight()` options on `svyset`; see [SVY] **svyset**.

The estimator for the poststratified total is

$$\widehat{Y}^P = \sum_{k=1}^{L_P} \frac{M_k}{\widehat{M}_k} \widehat{Y}_k = \sum_{k=1}^{L_P} \frac{M_k}{\widehat{M}_k} \sum_{j=1}^{m} I_{P_k}(j)\, w_j y_j$$

where

$$\widehat{M}_k = \sum_{j=1}^{m} I_{P_k}(j) w_j$$

The score variable for the poststratified total is

$$z_j(\widehat{Y}^P) = \sum_{k=1}^{L_P} I_{P_k}(j) \frac{M_k}{\widehat{M}_k} \left(y_j - \frac{\widehat{Y}_k}{\widehat{M}_k} \right)$$

Subpopulation estimation

Let S denote the set of sampled observations that belong to the subpopulation of interest, and define $I_S(j)$ to indicate if the jth observation falls within the subpopulation.

The estimator for the subpopulation total is

$$\widehat{Y}^S = \sum_{j=1}^{m} I_S(j)\, w_j y_j$$

and its score variable is

$$z_j(\widehat{Y}^S) = I_S(j)\, y_j$$

The estimator for the poststratified subpopulation total is

$$\widehat{Y}^{PS} = \sum_{k=1}^{L_P} \frac{M_k}{\widehat{M}_k} \widehat{Y}^S_k = \sum_{k=1}^{L_P} \frac{M_k}{\widehat{M}_k} \sum_{j=1}^{m} I_{P_k}(j) I_S(j)\, w_j y_j$$

and its score variable is

$$z_j(\widehat{Y}^{PS}) = \sum_{k=1}^{L_P} I_{P_k}(j) \frac{M_k}{\widehat{M}_k} \left\{ I_S(j)\, y_j - \frac{\widehat{Y}^S_k}{\widehat{M}_k} \right\}$$

References

Cochran, W. G. 1977. *Sampling Techniques*. 3rd ed. New York: Wiley.

Stuart, A., and J. K. Ord. 1994. *Kendall's Advanced Theory of Statistics: Distribution Theory, Vol I*. 6th ed. London: Arnold.

Also see

[R] **total postestimation** — Postestimation tools for total

[R] **mean** — Estimate means

[R] **proportion** — Estimate proportions

[R] **ratio** — Estimate ratios

[MI] **estimation** — Estimation commands for use with mi estimate

[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 estimation** — Estimation commands for survey data

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

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