total — Estimate totals

Description

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

Quick start

Total of continuous variable `v1`

```
total v1
```

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

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

Same 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*$_o$)	group over subpopulations defined by *varlist*$_o$
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
display_options	control column formats, line width, display of omitted variables and base and empty cells, and factor-variable labeling
coeflegend	display legend instead of statistics

varlist may contain factor variables; see **[U] 11.4.3 Factor variables**.

bayesboot, bootstrap, collect, 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*$_o$) specifies that estimates be computed for multiple subpopulations, which are identified by the different values of the variables in *varlist*$_o$. Only numeric, nonnegative, integer-valued variables are allowed in over(*varlist*$_o$).

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.

display_options: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofvlabel, fvwrap(#), fvwrapon(*style*), 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

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 https://www.stata-press.com/data/r19/total
(Fictional incidence of heart-attack data)
. total heartatk [pw=swgt], over(sex)
Total estimation                          Number of obs = 4,946
```

	Total	Std. err.	[95% conf.	interval]
c.heartatk@sex				
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(clustvar)	name of cluster variable
e(over)	*varlist* from over()
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

In addition to the above, the following is stored in r():

Matrices

r(table)	matrix containing the coefficients with their standard errors, test statistics, *p*-values, and confidence intervals

Note that results stored in r() are updated when the command is replayed and will be replaced when any r-class command is run after the estimation command.

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 $\hat{Y}$ for the population total Y is

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

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,

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

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

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

where

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

The score variable for the poststratified total is

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

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

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

and its score variable is

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

The estimator for the poststratified subpopulation total is

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

and its score variable is

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

References

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

Stuart, A., and J. K. Ord. 1994. *Distribution Theory*. Vol. 1 of *Kendall's Advanced Theory of Statistics*, 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**