**total** — Estimate totals

### Syntax

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

### options

- **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

### Description

**total** produces estimates of totals, along with standard errors.
Options

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.

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.

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:

coefflegend; 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 http://www.stata-press.com/data/r13/total
. total heartatk [pw=swgt], over(sex)

Total estimation                      Number of obs  =  4946

Male: sex = Male
Female: sex = Female

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>heartatk</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>944559</td>
<td>104372.3</td>
<td>739943</td>
</tr>
<tr>
<td>Female</td>
<td>581590</td>
<td>82855.59</td>
<td>419156.3</td>
</tr>
</tbody>
</table>

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_nameslist) 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
Survey data
The survey total estimator
The poststratified total estimator
Subpopulation estimation

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 \textit{iweight}), and if no weight is specified, define \( w_j = 1 \) for all \( j \). See the next section for \textit{pweight} data. The sum of the weights is an estimate of the population size:

\[
\hat{N} = \sum_{j=1}^{n} w_j
\]

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

\[
Y = \sum_{j=1}^{N} Y_j = N\bar{y}
\]

where \( \bar{y} \) is the population mean. The total is estimated as

\[
\hat{Y} = \hat{N}\bar{y}
\]

The variance estimator for the total is

\[
\hat{V}(\hat{Y}) = \hat{N}^2\hat{V}(\bar{y})
\]

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

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

\[
\hat{\text{Cov}}(\hat{X}, \hat{Y}) = \hat{N}^2\hat{\text{Cov}}(\bar{x}, \bar{y})
\]

where \( \hat{\text{Cov}}(\bar{x}, \bar{y}) \) 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 $j$th 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 $j$th 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

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

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 $j$th 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

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

where

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

The score variable for the poststratified total is

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

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

and its score variable is

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

References


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