mean — Estimate means

Description

mean produces estimates of means, along with standard errors.

Quick start

Mean, standard error, and 95% confidence interval for v1
mean v1

Also compute statistics for v2
mean v1 v2

As above, but for each level of categorical variable catvar1
mean v1 v2, over(catvar1)

Weighting by probability weight wvar
mean v1 v2 [pweight=wvar]

Population mean using svyset data
svy: mean v3

Subpopulation means for each level of categorical variable catvar2 using svyset data
svy: mean v3, over(catvar2)

Test equality of two subpopulation means
svy: mean v3, over(catvar2)
test v3@1.catvar2 = v3@2.catvar2

Menu

Statistics  >  Summaries, tables, and tests  >  Summary and descriptive statistics  >  Means
Syntax

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

<table>
<thead>
<tr>
<th>options</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td></td>
</tr>
<tr>
<td>stdize(varname)</td>
<td>variable identifying strata for standardization</td>
</tr>
<tr>
<td>stdweight(varname)</td>
<td>weight variable for standardization</td>
</tr>
<tr>
<td>nostdrescale</td>
<td>do not rescale the standard weight variable</td>
</tr>
<tr>
<td><strong>if/in/over</strong></td>
<td></td>
</tr>
<tr>
<td>over(varlist_o)</td>
<td>group over subpopulations defined by varlist_o</td>
</tr>
<tr>
<td><strong>SE/Cluster</strong></td>
<td></td>
</tr>
<tr>
<td>vce(vcetype)</td>
<td>vcetype may be analytic, cluster clustvar, bootstrap, or jackknife</td>
</tr>
<tr>
<td><strong>Reporting</strong></td>
<td></td>
</tr>
<tr>
<td>level(#)</td>
<td>set confidence level; default is level(95)</td>
</tr>
<tr>
<td>noheader</td>
<td>suppress table header</td>
</tr>
<tr>
<td>display_options</td>
<td>control column formats, line width, display of omitted variables and base and empty cells, and factor-variable labeling</td>
</tr>
<tr>
<td>coeflegend</td>
<td>display legend instead of statistics</td>
</tr>
</tbody>
</table>

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

Weights are not allowed with the bootstrap prefix; see [R] bootstrap.

vce() and weights are not allowed with the svy prefix; see [SVY] svy.

See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.

**Options**

- **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. `nostdrescale` prevents the standardization weights from being rescaled within the `over()` groups. This option requires `stdize()` but is ignored if the `over()` option is not specified.

- **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)`.
**Remarks and examples**

**Example 1**

Using the fuel data from example 3 of [R] *ttest*, we estimate the average mileage of the cars without the fuel treatment (mpg1) and those with the fuel treatment (mpg2).

```plaintext
use https://www.stata-press.com/data/r16/fuel
mean mpg1 mpg2
```

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>mpg1</td>
<td>21</td>
<td>.7881701</td>
<td>19.26525</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>22.73475</td>
</tr>
<tr>
<td>mpg2</td>
<td>22.75</td>
<td>.9384465</td>
<td>20.68449</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>24.81551</td>
</tr>
</tbody>
</table>

Using these results, we can test the equality of the mileage between the two groups of cars.

```plaintext
test mpg1 = mpg2
( 1)  mpg1 - mpg2 = 0
F( 1, 11) = 5.04
Prob > F = 0.0463
```
Example 2

In example 1, the joint observations of mpg1 and mpg2 were used to estimate a covariance between their means.

```stata
. matrix list e(V)
symmetric e(V)[2,2]
    mpg1   mpg2
  mpg1  .62121212
  mpg2  .4469697  .88068182
```

If the data were organized this way out of convenience but the two variables represent independent samples of cars (coincidentally of the same sample size), we should reshape the data and use the `over()` option to ensure that the covariance between the means is zero.

```stata
. use https://www.stata-press.com/data/r16/fuel
. stack mpg1 mpg2, into(mpg) clear
. rename _stack trt
. label define trt_lab 1 "without" 2 "with"
. label values trt trt_lab
. label var trt "Fuel treatment"
. mean mpg, over(trt)
Mean estimation
Number of obs  =  24

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
</table>
| c.mpg@trt
  without  | 21.7881701 | 19.36955 | 22.63045 |
  with     | 22.75 | .9384465 | 20.80868  | 24.69132 |

. matrix list e(V)
symmetric e(V)[2,2]
    c.mpg@  c.mpg@
  1.trt  2.trt
      .62121212
  c.mpg@1.trt  .88068182
  c.mpg@2.trt  0
```

Now, we can test the equality of the mileage between the two independent groups of cars.

```stata
. test mpg@1.trt = mpg@2.trt
(  1)  c.mpg@1.trt - c.mpg@2.trt = 0
F(   1, 23) = 2.04
Prob > F = 0.1667
```
Example 3: Standardized Means

Suppose that we collected the blood pressure data from example 2 of [R] dstdize, and we wish to obtain standardized high blood pressure rates for each city in 1990 and 1992, using, as the standard, the age, sex, and race distribution of the four cities and two years combined. Our rate is really the mean of a variable that indicates whether a sampled individual has high blood pressure. First, we generate the strata and weight variables from our standard distribution, and then use mean to compute the rates.

```stata
use https://www.stata-press.com/data/r16/hbp, clear
gen strata = group(age race sex) if inlist(year, 1990, 1992)
(by strata, sort: gen stdw = _N)
.mean hbp, over(city year) stdize(strata) stdweight(stdw)
```

Mean estimation

<table>
<thead>
<tr>
<th>c.hbp@city#year</th>
<th>Mean</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1990</td>
<td>.058642</td>
<td>.0296273</td>
<td>.0004182 .1168657</td>
</tr>
<tr>
<td>1 1992</td>
<td>.011765</td>
<td>.0113187</td>
<td>-.0104789 .0340083</td>
</tr>
<tr>
<td>2 1990</td>
<td>.048872</td>
<td>.0238958</td>
<td>.0019121 .0958322</td>
</tr>
<tr>
<td>2 1992</td>
<td>.014574</td>
<td>.007342</td>
<td>.0001455 .0290025</td>
</tr>
<tr>
<td>3 1990</td>
<td>.101121</td>
<td>.0268566</td>
<td>.0483425 .1538998</td>
</tr>
<tr>
<td>3 1992</td>
<td>.081057</td>
<td>.0227021</td>
<td>.0364435 .1256719</td>
</tr>
<tr>
<td>5 1990</td>
<td>.027778</td>
<td>.0155121</td>
<td>-.0027066 .0582622</td>
</tr>
<tr>
<td>5 1992</td>
<td>.054893</td>
<td>0</td>
<td>.   .</td>
</tr>
</tbody>
</table>

The standard error of the high blood pressure rate estimate is missing for city 5 in 1992 because there was only one individual with high blood pressure; that individual was the only person observed in the stratum of white males 30–35 years old.

By default, mean rescales the standard weights within the over() groups. In the following, we use the nostdrescale option to prevent this, thus reproducing the results in [R] dstdize.

```stata
.mean hbp, over(city year) stdize(strata) stdweight(stdw) nostdrescale
```

Mean estimation

<table>
<thead>
<tr>
<th>c.hbp@city#year</th>
<th>Mean</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1990</td>
<td>.007330</td>
<td>.0037034</td>
<td>.0000523 .0146082</td>
</tr>
<tr>
<td>1 1992</td>
<td>.001543</td>
<td>.0014847</td>
<td>-.0013745 .004461</td>
</tr>
<tr>
<td>2 1990</td>
<td>.007881</td>
<td>.0038536</td>
<td>.0003084 .0154544</td>
</tr>
<tr>
<td>2 1992</td>
<td>.002507</td>
<td>.0012633</td>
<td>.000025 .0049904</td>
</tr>
<tr>
<td>3 1990</td>
<td>.155271</td>
<td>.0041238</td>
<td>.007423 .0236312</td>
</tr>
<tr>
<td>3 1992</td>
<td>.008130</td>
<td>.0027772</td>
<td>.0036556 .012606</td>
</tr>
<tr>
<td>5 1990</td>
<td>.003922</td>
<td>.0021904</td>
<td>-.0003822 .0082268</td>
</tr>
<tr>
<td>5 1992</td>
<td>.008873</td>
<td>0</td>
<td>.   .</td>
</tr>
</tbody>
</table>
Example 4: profile plots and contrasts

The first example in [R] marginsplot shows how to use margins and marginsplot to get profile plots from a linear regression. We can similarly explore the data using marginsplot after mean with the over() option. Here we use marginsplot to plot the means of systolic blood pressure for each age group.

```
. use https://www.stata-press.com/data/r16/nhanes2, clear
. mean bpsystol, over(agegrp)
```

```
<table>
<thead>
<tr>
<th>agegrp</th>
<th>Mean</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-29</td>
<td>117.3466</td>
<td>.3247329</td>
<td>116.71 117.9831</td>
</tr>
<tr>
<td>30-39</td>
<td>120.2374</td>
<td>.4095845</td>
<td>119.4345 121.0402</td>
</tr>
<tr>
<td>40-49</td>
<td>126.9442</td>
<td>.532033</td>
<td>125.9013 127.9871</td>
</tr>
<tr>
<td>50-59</td>
<td>135.6754</td>
<td>.6061842</td>
<td>134.4872 136.8637</td>
</tr>
<tr>
<td>60-69</td>
<td>141.5227</td>
<td>.4433527</td>
<td>140.6537 142.3918</td>
</tr>
<tr>
<td>70+</td>
<td>148.1765</td>
<td>.8321116</td>
<td>146.5454 149.8076</td>
</tr>
</tbody>
</table>
```

```
. marginsplot
```

Variables that uniquely identify means: agegrp

We see that the mean systolic blood pressure increases with age. We can use contrast to formally test whether each mean is different from the mean in the previous age group using the ar. contrast operator; see [R] contrast for more information on this command.
The first row of the output reports that the mean systolic blood pressure for the 30–39 age group is 2.89 higher than the mean for the 20–29 age group. The mean for the 40–49 age group is 6.71 higher than the mean for the 30–39 age group, and so on. Each of these differences is significantly different from zero.

We can include both `agegrp` and `sex` in the `over()` option to estimate means separately for men and women in each age group.

```stata
. mean bpsystol, over(agegrp sex)
```

### Mean estimation

<table>
<thead>
<tr>
<th>c.bpsystol@agegrp#sex</th>
<th>Mean</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-29#Male</td>
<td>123.8862</td>
<td>0.4528516</td>
<td>122.9985 124.7739</td>
</tr>
<tr>
<td>20-29#Female</td>
<td>111.2849</td>
<td>0.3898972</td>
<td>110.5206 112.0492</td>
</tr>
<tr>
<td>30-39#Male</td>
<td>124.6818</td>
<td>0.5619855</td>
<td>123.5802 125.7834</td>
</tr>
<tr>
<td>30-39#Female</td>
<td>116.2207</td>
<td>0.5572103</td>
<td>115.1284 117.3129</td>
</tr>
<tr>
<td>40-49#Male</td>
<td>129.0033</td>
<td>0.7080788</td>
<td>127.6153 130.3912</td>
</tr>
<tr>
<td>40-49#Female</td>
<td>125.0468</td>
<td>0.7802558</td>
<td>123.5174 126.5763</td>
</tr>
<tr>
<td>50-59#Male</td>
<td>136.0864</td>
<td>0.855435</td>
<td>134.4096 137.7632</td>
</tr>
<tr>
<td>50-59#Female</td>
<td>135.3164</td>
<td>0.8356015</td>
<td>133.6393 136.9935</td>
</tr>
<tr>
<td>60-69#Male</td>
<td>140.7451</td>
<td>0.6059786</td>
<td>139.5572 141.9312</td>
</tr>
<tr>
<td>60-69#Female</td>
<td>142.2368</td>
<td>0.6427981</td>
<td>140.9767 143.4968</td>
</tr>
<tr>
<td>70+#Male</td>
<td>146.3951</td>
<td>1.141126</td>
<td>144.1583 148.6319</td>
</tr>
<tr>
<td>70+#Female</td>
<td>149.6599</td>
<td>1.189975</td>
<td>147.3273 151.9924</td>
</tr>
</tbody>
</table>
. marginsplot
Variables that uniquely identify means: agegrp sex

Estimated Means of bpsystol with 95% CIs

Are the means different for men and women within each age group? We can again perform the tests using contrast. This time, we will use `r.sex` to obtain contrasts comparing men and women and use `@agegrp` to request that the tests are performed for each age group.

. contrast r.sex#c.bpsystol@agegrp, effects nowald
Contrasts of means

| Contrast | Std. Err. | t    | P>|t|  | [95% Conf. Interval] |
|----------|-----------|------|------|---------------------|
| sex@agegrp#c.bpsystol (Female vs Male) 20-29 | -12.60132 | 0.5975738 | -21.09 | 0.000 | -13.77268 -11.42996 |
| (Female vs Male) 30-39 | -8.461161 | 0.7913981 | -10.69 | 0.000 | -10.01245 -6.909868 |
| (Female vs Male) 40-49 | -3.956451 | 1.053648 | -3.76 | 0.000 | -6.021805 -1.891097 |
| (Female vs Male) 50-59 | -0.7699782 | 1.209886 | -0.64 | 0.525 | -3.141588 1.601631 |
| (Female vs Male) 60-69 | 1.491684 | 0.8834022 | 1.69 | 0.091 | -0.2399545 3.223323 |
| (Female vs Male) 70+ | 3.264762 | 1.648699 | 1.98 | 0.048 | 0.0329927 6.496531 |
Using a 0.05 significance level, we find that the mean systolic blood pressure is different for men and women in all age groups except the fifties and sixties.

**Video example**

Descriptive statistics in Stata

**Stored results**

`mean` stores the following in `e()`:

Scalars

- `e(N)` number of observations
- `e(N_over)` number of subpopulations
- `e(N_stdize)` number of standard strata
- `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)` `mean`
- `e(cmdline)` command as typed
- `e(varlist)` `varlist`
- `e(stdize)` `varname` from `stdize()`
- `e(stdweight)` `varname` from `stdweight()`
- `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 mean estimates
- `e(V)` (co)variance estimates
- `e(sd)` vector of standard deviation estimates
- `e(N)` vector of numbers of nonmissing observations
- `e(N_stdsum)` number of nonmissing observations within the standard strata
- `e(p_stdize)` standardizing proportions
- `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 mean estimator
- Survey data
- The survey mean estimator
- The standardized mean estimator
- The poststratified mean estimator
- The standardized poststratified mean estimator
- Subpopulation estimation
The mean estimator

Let $y$ be the variable on which we want to calculate the mean and $y_j$ an individual observation on $y$, where $j = 1, \ldots, n$ and $n$ is the sample size. Let $w_j$ be the weight, and if no weight is specified, define $w_j = 1$ for all $j$. For aweights, the $w_j$ are normalized to sum to $n$. See The survey mean estimator for pweighted data.

Let $W$ be the sum of the weights

$$W = \sum_{j=1}^{n} w_j$$

The mean is defined as

$$\bar{y} = \frac{1}{W} \sum_{j=1}^{n} w_j y_j$$

The default variance estimator for the mean is

$$\hat{V}(\bar{y}) = \frac{1}{W(W-1)} \sum_{j=1}^{n} w_j (y_j - \bar{y})^2$$

The standard error of the mean is the square root of the variance.

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

$$\hat{\text{Cov}}(\bar{x}, \bar{y}) = \frac{1}{W(W-1)} \sum_{j=1}^{n} w_j (x_j - \bar{x})(y_j - \bar{y})$$

Survey data

See [SVY] Variance estimation, [SVY] Direct standardization, and [SVY] Poststratification for discussions that provide background information for the following formulas. The following formulas are derived from the fact that the mean is a special case of the ratio estimator where the denominator variable is one, $x_j = 1$; see [R] ratio.

The survey mean 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 mean for the item of interest is $\bar{Y} = Y/M$ where

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

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 for the mean is \( \bar{y} = \hat{Y} / \hat{M} \), where

\[
\hat{Y} = \sum_{j=1}^{m} w_j y_j \quad \text{and} \quad \hat{M} = \sum_{j=1}^{m} w_j
\]

and \( w_j \) is a sampling weight. The score variable for the mean estimator is

\[
z_j(\bar{y}) = \frac{y_j - \bar{y}}{M} = \frac{\hat{M}y_j - \hat{Y}}{\hat{M}^2}
\]

The standardized mean estimator

Let \( D_g \) denote the set of sampled observations that belong to the \( g \)th standard stratum and define \( I_{D_g}(j) \) to indicate if the \( j \)th observation is a member of the \( g \)th standard stratum; where \( g = 1, \ldots, L_D \) and \( L_D \) is the number of standard strata. Also, let \( \pi_g \) denote the fraction of the population that belongs to the \( g \)th standard stratum, thus \( \pi_1 + \cdots + \pi_{L_D} = 1 \). \( \pi_g \) is derived from the \texttt{stdweight()} option.

The estimator for the standardized mean is

\[
\bar{y}^D = \sum_{g=1}^{L_D} \pi_g \frac{\hat{Y}_g}{\hat{M}_g}
\]

where

\[
\hat{Y}_g = \sum_{j=1}^{m} I_{D_g}(j) w_j y_j \quad \text{and} \quad \hat{M}_g = \sum_{j=1}^{m} I_{D_g}(j) w_j
\]

The score variable for the standardized mean is

\[
z_j(\bar{y}^D) = \sum_{g=1}^{L_D} \pi_g I_{D_g}(j) \frac{\hat{M}_g y_j - \hat{Y}_g}{\hat{M}_g^2}
\]

The poststratified mean 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 \texttt{poststrata()} and \texttt{postweight()} options on \texttt{svyset}; see \texttt{[SVY] svyset}.

The estimator for the poststratified mean is

\[
\bar{y}^P = \frac{\hat{Y}^P}{\hat{M}^P} = \frac{\hat{Y}^P}{\hat{M}}
\]

where

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

The score variable for the poststratified mean is
\[
z_j(\hat{y}_P) = \frac{z_j(\hat{Y}_P)}{M} = \frac{1}{M} \sum_{k=1}^{L_P} I_{P_k}(j) \frac{M_k}{\hat{M}_k} \left( y_j - \frac{\hat{Y}_k}{\hat{M}_k} \right)
\]

The standardized poststratified mean estimator

The estimator for the standardized poststratified mean is
\[
\bar{y}^{DP} = \sum_{g=1}^{L_D} \pi_g \frac{\hat{Y}_P g}{\hat{M}_P g}
\]
where
\[
\hat{Y}_P g = \sum_{k=1}^{L_P} \frac{M_k}{\hat{M}_k} \hat{Y}_{g,k} = \sum_{k=1}^{L_P} \frac{M_k}{\hat{M}_k} \sum_{j=1}^{m} I_{D_g}(j) I_{P_k}(j) w_j y_j
\]
and
\[
\hat{M}_P g = \sum_{k=1}^{L_P} \frac{M_k}{\hat{M}_k} \hat{M}_{g,k} = \sum_{k=1}^{L_P} \frac{M_k}{\hat{M}_k} \sum_{j=1}^{m} I_{D_g}(j) I_{P_k}(j) w_j
\]

The score variable for the standardized poststratified mean is
\[
z_j(\bar{y}^{DP}) = \sum_{g=1}^{L_D} \pi_g \frac{\hat{M}_P g}{\hat{M}_P g} \left( \hat{Y}_P g - \hat{M}_P g \right)
\]
where
\[
z_j(\hat{Y}_P) = \sum_{k=1}^{L_P} I_{P_k}(j) \frac{M_k}{\hat{M}_k} \left\{ I_{D_g}(j) y_j - \frac{\hat{Y}_{g,k}}{\hat{M}_k} \right\}
\]
and
\[
z_j(\hat{M}_P g) = \sum_{k=1}^{L_P} I_{P_k}(j) \frac{M_k}{\hat{M}_k} \left\{ I_{D_g}(j) - \frac{\hat{M}_{g,k}}{\hat{M}_k} \right\}
\]

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 mean is \( \hat{y}^S = \frac{\hat{Y}^S}{\hat{M}^S} \), where

\[
\hat{Y}^S = \sum_{j=1}^{m} I_S(j) w_j y_j \quad \text{and} \quad \hat{M}^S = \sum_{j=1}^{m} I_S(j) w_j
\]

Its score variable is

\[
z_j(\hat{y}^S) = I_S(j) \frac{y_j - \hat{y}^S}{\hat{M}^S} = I_S(j) \frac{\hat{M}^S y_j - \hat{Y}^S}{(\hat{M}^S)^2}
\]

The estimator for the standardized subpopulation mean is

\[
\bar{y}^{DS} = \sum_{g=1}^{L_D} \pi_g \frac{\hat{Y}_g^S}{\hat{M}_g^S}
\]

where

\[
\hat{Y}_g^S = \sum_{j=1}^{m} I_{D_g}(j) I_S(j) w_j y_j \quad \text{and} \quad \hat{M}_g^S = \sum_{j=1}^{m} I_{D_g}(j) I_S(j) w_j
\]

Its score variable is

\[
z_j(\bar{y}^{DS}) = \sum_{g=1}^{L_D} \pi_g I_{D_g}(j) I_S(j) \frac{\hat{M}_g^S y_j - \hat{Y}_g^S}{(\hat{M}_g^S)^2}
\]

The estimator for the poststratified subpopulation mean is

\[
\bar{y}^{PS} = \frac{\hat{Y}^{PS}}{\hat{M}^{PS}}
\]

where

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

and

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

Its score variable is

\[
z_j(\bar{y}^{PS}) = \frac{\hat{M}^{PS} z_j(\hat{Y}^{PS}) - \hat{Y}^{PS} z_j(\hat{M}^{PS})}{(\hat{M}^{PS})^2}
\]

where

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

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

The estimator for the standardized poststratified subpopulation mean is

\[ \bar{y}^{DPS} = \sum_{g=1}^{L_D} \pi_g \frac{\hat{Y}^{PS}_g}{M^{PS}_g} \]

where

\[ \hat{Y}^{PS}_g = \sum_{k=1}^{L_P} \frac{M_k}{M_k} \hat{Y}^{S}_{g,k} = \sum_{k=1}^{L_P} \frac{M_k}{M_k} \sum_{j=1}^{m} I_{D_g}(j) I_{P_k}(j) I_S(j) w_j y_j \]

and

\[ \hat{M}^{PS}_g = \sum_{k=1}^{L_P} \frac{M_k}{M_k} \hat{M}^{S}_{g,k} = \sum_{k=1}^{L_P} \frac{M_k}{M_k} \sum_{j=1}^{m} I_{D_g}(j) I_{P_k}(j) I_S(j) w_j \]

Its score variable is

\[ z_j(\bar{y}^{DPS}) = \sum_{g=1}^{L_D} \pi_g \frac{\hat{M}^{PS}_g z_j(\hat{Y}^{PS}_g) - \hat{Y}^{PS}_g z_j(\hat{M}^{PS}_g)}{(\hat{M}^{PS}_g)^2} \]

where

\[ z_j(\hat{Y}^{PS}_g) = \sum_{k=1}^{L_P} I_{P_k}(j) \frac{M_k}{\hat{M}_k} \left\{ I_{D_g}(j) I_S(j) y_j - \frac{\hat{Y}^{S}_{g,k}}{\hat{M}_k} \right\} \]

and

\[ z_j(\hat{M}^{PS}_g) = \sum_{k=1}^{L_P} I_{P_k}(j) \frac{M_k}{\hat{M}_k} \left\{ I_{D_g}(j) I_S(j) - \frac{\hat{M}^{S}_{g,k}}{\hat{M}_k} \right\} \]

References


Also see

[R] mean postestimation — Postestimation tools for mean
[R] ameans — Arithmetic, geometric, and harmonic means
[R] proportion — Estimate proportions
[R] ratio — Estimate ratios
[R] summarize — Summary statistics
[R] total — Estimate totals

[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