

**ratio** — Estimate ratios

[Description](#)  
[Options](#)  
[References](#)

[Quick start](#)  
[Remarks and examples](#)  
[Also see](#)

[Menu](#)  
[Stored results](#)

[Syntax](#)  
[Methods and formulas](#)

## Description

`ratio` produces estimates of ratios, along with standard errors.

## Quick start

Estimate, standard error, and 95% confidence interval for the ratio of `v1` to `v2`

```
ratio v1/v2
```

With bootstrap standard errors

```
ratio v1/v2, vce(bootstrap)
```

Ratios of `v1` to `v2` and `v3` to `v2`

```
ratio (v1/v2) (v3/v2)
```

As above, but name the ratios `ratio1` and `ratio2`

```
ratio (ratio1: v1/v2) (ratio2: v3/v2)
```

Test that `ratio1` is equal to `ratio2`

```
test ratio1 = ratio2
```

Ratio of `v1` to `v2` over strata defined by levels of `svar`

```
ratio v1/v2, over(svar)
```

Direct standardization across categories `cvar`, weighting by standardization weight `wvar`

```
ratio v1/v2, stdize(cvar) stdweight(wvar)
```

## Menu

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

## Syntax

### Basic syntax

```
ratio [name:] varname [/] varname
```

### Full syntax

```
ratio ([name:] varname [/] varname)
      ([(name:] varname [/] varname) ...] [if] [in] [weight] [, options]
```

<i>options</i>	Description
<b>Model</b>	
<code>stdize(<i>varname</i>)</code>	variable identifying strata for standardization
<code>stdweight(<i>varname</i>)</code>	weight variable for standardization
<code>nostdrescale</code>	do not rescale the standard weight variable
<b>if/in/over</b>	
<code>over(<i>varlist</i> [, <u>no</u>label])</code>	group over subpopulations defined by <i>varlist</i> ; optionally, suppress group labels
<b>SE/Cluster</b>	
<code>vce(<i>vcetype</i>)</code>	<i>vcetype</i> may be <u>linearized</u> , <u>cluster</u> <i>clustvar</i> , <u>bootstrap</u> , or <u>jackknife</u>
<b>Reporting</b>	
<code>level(#)</code>	set confidence level; default is level(95)
<code>noheader</code>	suppress table header
<code>nolegend</code>	suppress table legend
<code>display_options</code>	control column formats and line width
<code>coeflegend</code>	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; see [MI] mi estimate.	
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

### 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 [ , 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.

`no`label 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 (`linearized`), that allow for intragroup correlation (`cluster clustvar`), and that use bootstrap or jackknife methods (`bootstrap`, `jackknife`); see [R] [vce\\_option](#).

`vce(linearized)`, the default, uses the linearized or sandwich estimator of variance.

## Reporting

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

`no`header prevents the table header from being displayed. This option implies `no`legend.

`no`legend prevents the table legend identifying the subpopulations from being displayed.

*display\_options*: `cformat(%fmt)` and `no`lstretch; see [R] [estimation options](#).

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

`co`eflegend; see [R] [estimation options](#).

## Remarks and examples

## ▷ Example 1

Using the fuel data from [example 3](#) of [R] `ttest`, we estimate the ratio of mileage for the cars without the fuel treatment (`mpg1`) to those with the fuel treatment (`mpg2`).

```
. use http://www.stata-press.com/data/r14/fuel
. ratio myratio: mpg1/mpg2
Ratio estimation          Number of obs   =          12
    myratio: mpg1/mpg2
```

	Linearized			
	Ratio	Std. Err.	[95% Conf. Interval]	
myratio	.9230769	.032493	.8515603	.9945936

Using these results, we can test to see if this ratio is significantly different from one.

```
. test _b[myratio] = 1
( 1) myratio = 1
      F( 1, 11) =    5.60
      Prob > F =    0.0373
```

We find that the ratio is different from one at the 5% significance level but not at the 1% significance level.



## ▷ Example 2

Using state-level census data, we want to test whether the marriage rate is equal to the death rate.

```
. use http://www.stata-press.com/data/r14/census2
(1980 Census data by state)
. ratio (deathrate: death/pop) (marrate: marriage/pop)
Ratio estimation          Number of obs   =          50
    deathrate: death/pop
    marrate: marriage/pop
```

	Linearized			
	Ratio	Std. Err.	[95% Conf. Interval]	
deathrate	.0087368	.0002052	.0083244	.0091492
marrate	.0105577	.0006184	.009315	.0118005

```
. test _b[deathrate] = _b[marrate]
( 1) deathrate - marrate = 0
      F( 1, 49) =    6.93
      Prob > F =    0.0113
```



## Stored results

ratio stores the following in `e()`:

### Scalars

<code>e(N)</code>	number of observations
<code>e(N_over)</code>	number of subpopulations
<code>e(N_stdize)</code>	number of standard strata
<code>e(N_clust)</code>	number of clusters
<code>e(k_eq)</code>	number of equations in <code>e(b)</code>
<code>e(df_r)</code>	sample degrees of freedom
<code>e(rank)</code>	rank of <code>e(V)</code>

### Macros

<code>e(cmd)</code>	ratio
<code>e(cmdline)</code>	command as typed
<code>e(varlist)</code>	<i>varlist</i>
<code>e(stdize)</code>	<i>varname</i> from <code>stdize()</code>
<code>e(stdweight)</code>	<i>varname</i> from <code>stdweight()</code>
<code>e(wtype)</code>	weight type
<code>e(wexp)</code>	weight expression
<code>e(title)</code>	title in estimation output
<code>e(cluster)</code>	name of cluster variable
<code>e(over)</code>	<i>varlist</i> from <code>over()</code>
<code>e(over_labels)</code>	labels from <code>over()</code> variables
<code>e(over_namelist)</code>	names from <code>e(over_labels)</code>
<code>e(namelist)</code>	ratio identifiers
<code>e(vce)</code>	<i>vcetype</i> specified in <code>vce()</code>
<code>e(vcetype)</code>	title used to label Std. Err.
<code>e(properties)</code>	<code>b V</code>
<code>e(estat_cmd)</code>	program used to implement <code>estat</code>
<code>e(marginsnotok)</code>	predictions disallowed by <code>margins</code>

### Matrices

<code>e(b)</code>	vector of ratio estimates
<code>e(V)</code>	(co)variance estimates
<code>e(_N)</code>	vector of numbers of nonmissing observations
<code>e(_N_stdsum)</code>	number of nonmissing observations within the standard strata
<code>e(_p_stdize)</code>	standardizing proportions
<code>e(error)</code>	error code corresponding to <code>e(b)</code>

### Functions

<code>e(sample)</code>	marks estimation sample
------------------------	-------------------------

## Methods and formulas

Methods and formulas are presented under the following headings:

*The ratio estimator*  
*Survey data*  
*The survey ratio estimator*  
*The standardized ratio estimator*  
*The poststratified ratio estimator*  
*The standardized poststratified ratio estimator*  
*Subpopulation estimation*

## The ratio estimator

Let  $R = Y/X$  be the ratio to be estimated, where  $Y$  and  $X$  are totals; see [R] total. The estimate for  $R$  is  $\hat{R} = \hat{Y}/\hat{X}$  (the ratio of the sample totals). From the delta method (that is, a first-order Taylor expansion), the approximate variance of the sampling distribution of the linearized  $\hat{R}$  is

$$V(\hat{R}) \approx \frac{1}{\hat{X}^2} \left\{ V(\hat{Y}) - 2RCov(\hat{Y}, \hat{X}) + R^2V(\hat{X}) \right\}$$

Direct substitution of  $\hat{X}$ ,  $\hat{R}$ , and the estimated variances and covariance of  $\hat{X}$  and  $\hat{Y}$  leads to the following variance estimator:

$$\hat{V}(\hat{R}) = \frac{1}{\hat{X}^2} \left\{ \hat{V}(\hat{Y}) - 2\hat{R}\widehat{Cov}(\hat{Y}, \hat{X}) + \hat{R}^2\hat{V}(\hat{X}) \right\} \quad (1)$$

## Survey data

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

## The survey ratio estimator

Let  $Y_j$  and  $X_j$  be survey items for the  $j$ th individual in the population, where  $j = 1, \dots, M$  and  $M$  is the size of the population. The associated population ratio for the items of interest is  $R = Y/X$  where

$$Y = \sum_{j=1}^M Y_j \quad \text{and} \quad X = \sum_{j=1}^M X_j$$

Let  $y_j$  and  $x_j$  be the corresponding survey items for the  $j$ th sampled individual from the population, where  $j = 1, \dots, m$  and  $m$  is the number of observations in the sample.

The estimator  $\hat{R}$  for the population ratio  $R$  is  $\hat{R} = \hat{Y}/\hat{X}$ , where

$$\hat{Y} = \sum_{j=1}^m w_j y_j \quad \text{and} \quad \hat{X} = \sum_{j=1}^m w_j x_j$$

and  $w_j$  is a sampling weight. The score variable for the ratio estimator is

$$z_j(\hat{R}) = \frac{y_j - \hat{R}x_j}{\hat{X}} = \frac{\hat{X}y_j - \hat{Y}x_j}{\hat{X}^2}$$

## The standardized ratio 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  $g = 1, \dots, L_D$  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  $\pi_1 + \dots + \pi_{L_D} = 1$ . Note that  $\pi_g$  is derived from the `stdweight()` option.

The estimator for the standardized ratio is

$$\widehat{R}^D = \sum_{g=1}^{L_D} \pi_g \frac{\widehat{Y}_g}{\widehat{X}_g}$$

where

$$\widehat{Y}_g = \sum_{j=1}^m I_{D_g}(j) w_j y_j$$

and  $\widehat{X}_g$  is similarly defined. The score variable for the standardized ratio is

$$z_j(\widehat{R}^D) = \sum_{g=1}^{L_D} \pi_g I_{D_g}(j) \frac{\widehat{X}_g y_j - \widehat{Y}_g x_j}{\widehat{X}_g^2}$$

## The poststratified ratio 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  $k = 1, \dots, L_P$  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 ratio is

$$\widehat{R}^P = \frac{\widehat{Y}^P}{\widehat{X}^P}$$

where

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

and  $\widehat{X}^P$  is similarly defined. The score variable for the poststratified ratio is

$$z_j(\widehat{R}^P) = \frac{z_j(\widehat{Y}^P) - \widehat{R}^P z_j(\widehat{X}^P)}{\widehat{X}^P} = \frac{\widehat{X}^P z_j(\widehat{Y}^P) - \widehat{Y}^P z_j(\widehat{X}^P)}{(\widehat{X}^P)^2}$$

where

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

and  $z_j(\widehat{X}^P)$  is similarly defined.

## The standardized poststratified ratio estimator

The estimator for the standardized poststratified ratio is

$$\widehat{R}^{DP} = \sum_{g=1}^{L_D} \pi_g \frac{\widehat{Y}_g^P}{\widehat{X}_g^P}$$

where

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

and  $\widehat{X}_g^P$  is similarly defined. The score variable for the standardized poststratified ratio is

$$z_j(\widehat{R}^{DP}) = \sum_{g=1}^{L_D} \pi_g \frac{\widehat{X}_g^P z_j(\widehat{Y}_g^P) - \widehat{Y}_g^P z_j(\widehat{X}_g^P)}{(\widehat{X}_g^P)^2}$$

where

$$z_j(\widehat{Y}_g^P) = \sum_{k=1}^{L_p} I_{P_k}(j) \frac{M_k}{\widehat{M}_k} \left\{ I_{D_g}(j) y_j - \frac{\widehat{Y}_{g,k}}{\widehat{M}_k} \right\}$$

and  $z_j(\widehat{X}_g^P)$  is similarly defined.

## 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 ratio is  $\widehat{R}^S = \widehat{Y}^S / \widehat{X}^S$ , where

$$\widehat{Y}^S = \sum_{j=1}^m I_S(j) w_j y_j \quad \text{and} \quad \widehat{X}^S = \sum_{j=1}^m I_S(j) w_j x_j$$

Its score variable is

$$z_j(\widehat{R}^S) = I_S(j) \frac{y_j - \widehat{R}^S x_j}{\widehat{X}^S} = I_S(j) \frac{\widehat{X}^S y_j - \widehat{Y}^S x_j}{(\widehat{X}^S)^2}$$

The estimator for the standardized subpopulation ratio is

$$\widehat{R}^{DS} = \sum_{g=1}^{L_D} \pi_g \frac{\widehat{Y}_g^S}{\widehat{X}_g^S}$$

where

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

and  $\widehat{X}_g^S$  is similarly defined. Its score variable is

$$z_j(\widehat{R}^{DS}) = \sum_{g=1}^{L_D} \pi_g I_{D_g}(j) I_S(j) \frac{\widehat{X}_g^S y_j - \widehat{Y}_g^S x_j}{(\widehat{X}_g^S)^2}$$



The estimator for the poststratified subpopulation ratio is

$$\widehat{R}^{PS} = \frac{\widehat{Y}^{PS}}{\widehat{X}^{PS}}$$

where

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

and  $\widehat{X}^{PS}$  is similarly defined. Its score variable is

$$z_j(\widehat{R}^{PS}) = \frac{\widehat{X}^{PS} z_j(\widehat{Y}^{PS}) - \widehat{Y}^{PS} z_j(\widehat{X}^{PS})}{(\widehat{X}^{PS})^2}$$

where

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

and  $z_j(\widehat{X}^{PS})$  is similarly defined.

The estimator for the standardized poststratified subpopulation ratio is

$$\widehat{R}^{DPS} = \sum_{g=1}^{L_D} \pi_g \frac{\widehat{Y}_g^{PS}}{\widehat{X}_g^{PS}}$$

where

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

and  $\widehat{X}_g^{PS}$  is similarly defined. Its score variable is

$$z_j(\widehat{R}^{DPS}) = \sum_{g=1}^{L_D} \pi_g \frac{\widehat{X}_g^{PS} z_j(\widehat{Y}_g^{PS}) - \widehat{Y}_g^{PS} z_j(\widehat{X}_g^{PS})}{(\widehat{X}_g^{PS})^2}$$

where

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

and  $z_j(\widehat{X}_g^{PS})$  is similarly defined.

## 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] **ratio postestimation** — Postestimation tools for ratio

[R] **mean** — Estimate means

[R] **proportion** — Estimate proportions

[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**