**estat summarize** — Summarize estimation sample



## Description

`estat summarize` summarizes the variables used by the command and automatically restricts the sample to the estimation sample; it also summarizes the weight variable and cluster structure, if specified.

## Quick start

Summary statistics for all variables in the model using estimation sample

```
estat summarize
```

Add variable labels to output

```
estat summarize, labels
```

Obtain summary of estimation sample for each equation

```
estat summarize, equation
```

Ignore weights when calculating summary statistics after weighted estimation

```
estat summarize, noweights
```

## Menu for estat

Statistics > Postestimation

# Syntax

```
estat summarize [eqlist] [ , estat_summ_options ]
```

| *estat_summ_options* | Description |
|---|---|
| equation | display summary by equation |
| group | display summary by group; only after sem and gsem |
| labels | display variable labels |
| noheader | suppress the header |
| noweights | ignore weights |
| *display_options* | control row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling |

*eqlist* is rarely used and specifies the variables, with optional equation name, to be summarized. *eqlist* may be *varlist* or (*eqname*$_1$: *varlist*) (*eqname*$_2$: *varlist*) .... *varlist* may contain time-series operators; see **[U] 11.4.4 Time-series varlists**.

collect is allowed; see **[U] 11.1.10 Prefix commands**.

# Options

equation requests that the dependent variables and the independent variables in the equations be displayed in the equation-style format of estimation commands, repeating the summary information about variables entered in more than one equation.

group displays summary information separately for each group. group is only allowed after sem or gsem with a group() variable specified.

labels displays variable labels.

noheader suppresses the header.

noweights ignores the weights, if any, from the previous estimation command. The default when weights are present is to perform a weighted summarize on all variables except the weight variable itself. An unweighted summarize is performed on the weight variable.

*display_options*: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofvlabel, fvwrap(#), and fvwrapon(*style*); see [R] **Estimation options**.

# Remarks and examples

Often when fitting a model, you will also be interested in obtaining summary statistics, such as the sample means and standard deviations of the variables in the model. estat summarize makes this process simple. The output displayed is similar to that obtained by typing

```
. summarize varlist if e(sample)
```

without the need to type the *varlist* containing the dependent and independent variables.

## Example 1

Continuing with the example in [R] **estat ic**, here we summarize the variables by using `estat summarize`.

```
. use https://www.stata-press.com/data/r19/sysdsn1
(Health insurance data)
. mlogit insure age male nonwhite i.site
  (output omitted)
. estat summarize, noomitted
Estimation sample mlogit                 Number of obs =        615
```

| Variable | Mean | Std. dev. | Min | Max |
|---|---|---|---|---|
| insure | 1.596748 | .6225846 | 1 | 3 |
| age | 44.46832 | 14.18523 | 18.11087 | 86.07254 |
| male | .2504065 | .4335998 | 0 | 1 |
| nonwhite | .196748 | .3978638 | 0 | 1 |
| site | | | | |
| 2 | .3707317 | .4833939 | 0 | 1 |
| 3 | .3138211 | .4644224 | 0 | 1 |

The output in the previous example contains all the variables in one table, though `mlogit` presents its results in a multiple-equation format. For models in which the same variables appear in all equations, that is fine; but for other multiple-equation models, we may prefer to have the variables separated by the equation in which they appear. The `equation` option makes this possible.

### ▷ Example 2

Systems of simultaneous equations typically have different variables in each equation, and the `equation` option of `estat summarize` is helpful in such situations. In example 2 of [R] **reg3**, we have a model of supply and demand. We first refit the model and then call `estat summarize`.

```
. use https://www.stata-press.com/data/r19/supDem
. reg3 (Demand:quantity price pcompete income) (Supply:quantity price praw),
> endog(price)
```

(*output omitted*)

```
. estat summarize, equation
```

Estimation sample reg3 Number of obs = 49

| Variable | Mean | Std. dev. | Min | Max |
|---|---|---|---|---|
| depvar | | | | |
| quantity | 12.61818 | 2.774952 | 7.710694 | 20.04767 |
| quantity | 12.61818 | 2.774952 | 7.710694 | 20.04767 |
| demale | | | | |
| price | 32.70944 | 2.882684 | 26.38185 | 38.47692 |
| pcompete | 5.929975 | 3.508264 | .2076465 | 11.55491 |
| income | 7.811735 | 4.18859 | .5704173 | 14.00767 |
| Supply | | | | |
| price | 32.70944 | 2.882684 | 26.38185 | 38.47692 |
| praw | 4.740891 | 2.962565 | .1510276 | 9.79881 |

The first block of the table contains statistics on the dependent (or, more accurately, left-hand-side) variables, and because we specified quantity as the left-hand-side variable in both equations, it is listed twice. The second block refers to the variables in the first equation we specified, which we labeled "Demand" in our call to `reg3`; and the final block refers to the supply equation.

◁

## Stored results

`estat summarize` stores the following in `r()`:

Scalars

| | |
|---|---|
| `r(N_groups)` | number of groups (`group` only) |

Matrices

| | |
|---|---|
| `r(stats)` | $k \times 4$ matrix of means, standard deviations, minimums, and maximums |
| `r(stats[_#])` | $k \times 4$ matrix of means, standard deviations, minimums, and maximums for group #(`group` only) |

## Also see

[R] **estat** — Postestimation statistics

[R] **estat ic** — Display information criteria

[R] **estat vce** — Display covariance matrix estimates



For suggested citations, see the FAQ on citing Stata documentation.