**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

\texttt{estat summarize} \hspace{1mm} [\textit{eqlist}] \hspace{1mm} [, \textit{estat_summ_options}] \\

\textit{estat_summ_options} \hspace{1mm} Description

\begin{tabular}{|l|p{13cm}|}
\hline
\textit{equation} & display summary by equation \\
\textit{group} & display summary by group; only after \texttt{sem} and \texttt{gsem} \\
\textit{labels} & display variable labels \\
\textit{noheader} & suppress the header \\
\textit{noweights} & ignore weights \\
\textit{display_options} & control row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling \\
\hline
\end{tabular}

\textit{eqlist} is rarely used and specifies the variables, with optional equation name, to be summarized. \textit{eqlist} may be \textit{varlist} or \texttt{\textit{eqname1} : varlist} \texttt{\textit{eqname2} : varlist} \ldots \textit{varlist} may contain time-series operators; see [U] 11.4.4 Time-series \textit{varlists}.

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. \textit{group} is only allowed after \texttt{sem} or \texttt{gsem} with a \texttt{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 \texttt{summarize} on all variables except the weight variable itself. An unweighted \texttt{summarize} is performed on the weight variable.

display\_options: \texttt{nomitted}, \texttt{vsquish}, \texttt{noemptycells}, \texttt{baselabels}, \texttt{allbaselabels}, \texttt{nofvlabel}, \texttt{fwrap(\#)}, and \texttt{fwrapon(style)}; see [R] Estimation options.

Remarks and examples

\texttt{stata.com}

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. \texttt{estat summarize} makes this process simple. The output displayed is similar to that obtained by typing

\texttt{. summarize varlist if e(sample)}

without the need to type the \textit{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/r16/sysdsn1
(Health insurance data)
. mlogit insure age male nonwhite i.site
(output omitted)
. estat summarize, noomitted

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>insure</td>
<td>1.596748</td>
<td>.6225846</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>age</td>
<td>44.46832</td>
<td>14.18523</td>
<td>18.11087</td>
<td>86.07254</td>
</tr>
<tr>
<td>male</td>
<td>.2504065</td>
<td>.4335998</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>nonwhite</td>
<td>.196748</td>
<td>.3978638</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>site</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.3707317</td>
<td>.4833939</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>.3138211</td>
<td>.4644224</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

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/r16/supDem
.reg3 (Demand:quantity price pcompete income) (Supply:quantity price praw),
> endog(price)
(output omitted)
estat summarize, equation
```

<table>
<thead>
<tr>
<th>Estimation sample reg3</th>
<th>Number of obs = 49</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dependent Variable</strong></td>
<td><strong>Mean</strong></td>
</tr>
<tr>
<td>depvar</td>
<td></td>
</tr>
<tr>
<td>quantity</td>
<td>12.61818</td>
</tr>
<tr>
<td>quantity</td>
<td>12.61818</td>
</tr>
<tr>
<td><strong>Demand</strong></td>
<td></td>
</tr>
<tr>
<td>price</td>
<td>32.70944</td>
</tr>
<tr>
<td>pcompete</td>
<td>5.929975</td>
</tr>
<tr>
<td>income</td>
<td>7.811735</td>
</tr>
<tr>
<td><strong>Supply</strong></td>
<td></td>
</tr>
<tr>
<td>price</td>
<td>32.70944</td>
</tr>
<tr>
<td>praw</td>
<td>4.740891</td>
</tr>
</tbody>
</table>

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