estat summarize — Summarize estimation sample

Syntax

estat summarize [eqlist] [ , estat_summ_options ]

eqlist

estat_summ_options

equation display summary by equation

group display summary by group; only after sem

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

Syntax

eqlist is rarely used and specifies the variables, with optional equation name, to be summarized. eqlist may be

varlist or (eqname1: varlist) (eqname2: varlist) . . . varlist may contain time-series operators; see

[U] 11.4.4 Time-series varlists.

Menu for estat

Statistics > Postestimation > Reports and statistics

Description

estat summarize summarizes the variables used by the command and automatically restricts the

sample to e(sample); it also summarizes the weight variable and cluster structure, if specified.

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

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

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

\begin{verbatim}
. summarize varlist if e(sample)
\end{verbatim}

without the need to type the \texttt{varlist} containing the dependent and independent variables.

Example 1

Continuing with the example in [R] \texttt{estat ic}, here we summarize the variables by using \texttt{estat summarize}.

\begin{verbatim}
. use http://www.stata-press.com/data/r13/sysdsn1
(Health insurance data)
. mlogit insure age male nonwhite i.site
(output omitted)
. estat summarize, noomitted
\end{verbatim}

\begin{tabular}{l|ccc}
Variable & Mean & Std. Dev. & Min & Max \\
\hline
insure & 1.596748 & .6225846 & 1 & 3 \\
age & 44.46832 & 14.18523 & 18.1109 & 86.0725 \\
male & .2504065 & .4335998 & 0 & 1 \\
nonwhite & .196748 & .3978638 & 0 & 1 \\
site & & & & \\
2 & .3707317 & .4833939 & 0 & 1 \\
3 & .3138211 & .4644224 & 0 & 1 \\
\end{tabular}

The output in the previous example contains all the variables in one table, though \texttt{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 \texttt{equation} option makes this possible.

Example 2

Systems of simultaneous equations typically have different variables in each equation, and the \texttt{equation} option of \texttt{estat summarize} is helpful in such situations. In example 2 of [R] \texttt{reg3}, we have a model of supply and demand. We first refit the model and then call \texttt{estat summarize}.
. use http://www.stata-press.com/data/r13/supDem
. reg3 (Demand: quantity price pcompete income) (Supply: quantity price praw),
> endog(price)
(output omitted)
. estat summarize, equation

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>depvar</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>quantity</td>
<td>12.61818</td>
<td>2.774952</td>
<td>7.71069</td>
<td>20.0477</td>
</tr>
<tr>
<td>quantity</td>
<td>12.61818</td>
<td>2.774952</td>
<td>7.71069</td>
<td>20.0477</td>
</tr>
<tr>
<td>demale</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>price</td>
<td>32.70944</td>
<td>2.882684</td>
<td>26.3819</td>
<td>38.4769</td>
</tr>
<tr>
<td>pcompete</td>
<td>5.929975</td>
<td>3.508264</td>
<td>.207647</td>
<td>11.5549</td>
</tr>
<tr>
<td>income</td>
<td>7.811735</td>
<td>4.18859</td>
<td>.570417</td>
<td>14.0077</td>
</tr>
<tr>
<td>Supply</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>price</td>
<td>32.70944</td>
<td>2.882684</td>
<td>26.3819</td>
<td>38.4769</td>
</tr>
<tr>
<td>praw</td>
<td>4.740891</td>
<td>2.962565</td>
<td>.151028</td>
<td>9.79881</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 × 4 matrix of means, standard deviations, minimums, and maximums
r(stats[_#]) k × 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