**by — Repeat Stata command on subsets of the data**

**Description**

Most Stata commands allow the `by` prefix, which repeats the command for each group of observations for which the values of the variables in `varlist` are the same. `by` without the `sort` option requires that the data be sorted by `varlist`; see [D] sort.

Stata commands that work with the `by` prefix indicate this immediately following their syntax diagram by reporting, for example, “by is allowed; see [D] by” or “bootstrap, by, etc., are allowed; see [U] 11.1.10 Prefix commands”.

`by` and `bysort` are really the same command; `bysort` is just `by` with the `sort` option.

The `varlist1 (varlist2)` syntax is of special use to programmers. It verifies that the data are sorted by `varlist1 varlist2` and then performs a `by` as if only `varlist1` were specified. For instance,

```
by pid (time): generate growth = (bp - bp[_n-1])/bp
```

performs the `generate` by values of `pid` but first verifies that the data are sorted by `pid` and `time` within `pid`.

**Quick start**

Generate `newvar` as an observation number within each level of `catvar`

```
by catvar: generate newvar = _n
```

As above, but sort data by `catvar` first

```
by catvar, sort: generate newvar = _n
```

Same as above

```
bysort catvar: generate newvar = _n
```

As above, but sort by `v` within values of `catvar`

```
bysort catvar (v): generate newvar = _n
```

Generate `newvar` as an observation number for each observation in levels of `catvar` and `v`

```
bysort catvar v: generate newvar = _n
```

Note: Any command that accepts the `by` prefix may be substituted for `generate` above.
Syntax

\texttt{by \ varlist \ : \ stata\_cmd}

\texttt{bysort \ varlist \ : \ stata\_cmd}

The above diagrams show \texttt{by} and \texttt{bysort} as they are typically used. The full syntax of the commands is

\texttt{by \ varlist1 \ [(\ varlist2 \)] \[, \ sort \ rc0 \] : \ stata\_cmd}

\texttt{bysort \ varlist1 \ [(\ varlist2 \)] \[, \ rc0 \] : \ stata\_cmd}

Options

\texttt{sort} specifies that if the data are not already sorted by \texttt{varlist}, \texttt{by} should sort them.

\texttt{rc0} specifies that even if the \texttt{stata\_cmd} produces an error in one of the by-groups, then \texttt{by} is still to run the \texttt{stata\_cmd} on the remaining by-groups. The default action is to stop when an error occurs. \texttt{rc0} is especially useful when \texttt{stata\_cmd} is an estimation command and some by-groups have insufficient observations.

Remarks and examples

\texttt{stata.com}

\textbf{Example 1}

\begin{verbatim}
. use https://www.stata-press.com/data/r16/autornd
  (1978 Automobile Data)
. keep in 1/20
  (54 observations deleted)
. by mpg: egen mean_w = mean(weight)
  not sorted
  r(5);
. sort mpg
. by mpg: egen mean_w = mean(weight)
\end{verbatim}
by requires that the data be sorted. In the above example, we could have typed by mpg, sort: egen mean_w = mean(weight) or bysort mpg: egen mean_w = mean(weight) rather than the separate sort; all would yield the same results.

For more examples, see [U] 11.1.2 by varlist:, [U] 11.5 by varlist: construct, and [U] 13.7 Explicit subscripting. For extended introductions with detailed examples, see Cox (2002) and Mitchell (2020, chap. 8).

Technical note

by repeats the stata_cmd for each group defined by varlist. If stata_cmd stores results, only the results from the last group on which stata_cmd executes will be stored.

References

Also see

[D] sort — Sort data
[D] statsby — Collect statistics for a command across a by list
[P] byable — Make programs byable
[P] foreach — Loop over items
[P] forvalues — Loop over consecutive values
[P] while — Looping
[U] 11.1.2 by varlist:
[U] 11.1.10 Prefix commands
[U] 11.4 varname and varlists
[U] 11.5 by varlist: construct