cluster utility — List, rename, use, and drop cluster analyses

Description

These cluster utility commands allow you to view and manipulate the cluster objects that you have created. See [MV] cluster for an overview of cluster analysis and for the available cluster commands. If you want even more control over your cluster objects, or if you are programming new cluster subprograms, more cluster programmer utilities are available; see [MV] cluster programming utilities for details.

The cluster dir command provides a directory-style listing of all the currently defined clusters. The cluster list command provides a detailed listing of the specified clusters or of all current clusters if no cluster names are specified. The default action is to list all the information attached to the clusters. You may limit the type of information listed by specifying particular options.

The cluster drop command removes the named clusters. The keyword _all specifies that all current cluster analyses be dropped.

Stata cluster analyses are referred to by name. Many cluster commands default to using the most recently defined cluster analysis if no cluster name is provided. The cluster use command sets the specified cluster analysis as the most recently executed cluster analysis, so that, by default, this cluster analysis will be used if the cluster name is omitted from many of the cluster commands. You may use the * and ? name-matching characters to shorten the typing of cluster names; see [U] 11.2 Abbreviation rules.

cluster rename allows you to rename a cluster analysis without changing any of the variable names attached to the cluster analysis. The cluster renamevar command, on the other hand, allows you to rename the variables attached to a cluster analysis and to update the cluster object with the new variable names. Do not use the rename command (see [D] rename) to rename variables attached to a cluster analysis because this would invalidate the cluster object. Use the cluster renamevar command instead.

Menu

cluster list
Statistics > Multivariate analysis > Cluster analysis > Postclustering > Detailed listing of clusters

cluster drop
Statistics > Multivariate analysis > Cluster analysis > Postclustering > Drop cluster analyses

cluster rename
Statistics > Multivariate analysis > Cluster analysis > Postclustering > Rename a cluster or cluster variables
2 cluster utility — List, rename, use, and drop cluster analyses

Syntax

Directory-style listing of currently defined clusters

    cluster dir

Detailed listing of clusters

    cluster list [clnamelist] [, list_options]

Drop cluster analyses

    cluster drop {clnamelist | _all}

Mark a cluster analysis as the most recent one

    cluster use clname

Rename a cluster

    cluster rename oldclname newclname

Rename variables attached to a cluster

    cluster renamevar oldvarname newvar [, name(clname)]

    cluster renamevar oldstub newstub, prefix [name(clname)]

list_options Description

<table>
<thead>
<tr>
<th>Options</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>notes</td>
<td>list cluster notes</td>
</tr>
<tr>
<td>type</td>
<td>list cluster analysis type</td>
</tr>
<tr>
<td>method</td>
<td>list cluster analysis method</td>
</tr>
<tr>
<td>dissimilarity</td>
<td>list cluster analysis dissimilarity measure</td>
</tr>
<tr>
<td>similarity</td>
<td>list cluster analysis similarity measure</td>
</tr>
<tr>
<td>vars</td>
<td>list variable names attached to the cluster analysis</td>
</tr>
<tr>
<td>chars</td>
<td>list any characteristics attached to the cluster analysis</td>
</tr>
<tr>
<td>other</td>
<td>list any “other” information</td>
</tr>
<tr>
<td>all</td>
<td>list all items and information attached to the cluster; the default</td>
</tr>
</tbody>
</table>

all does not appear in the dialog box.

Options for cluster list

<table>
<thead>
<tr>
<th>Options</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>notes</td>
<td>specifies that cluster notes be listed.</td>
</tr>
<tr>
<td>type</td>
<td>specifies that the type of cluster analysis be listed.</td>
</tr>
<tr>
<td>method</td>
<td>specifies that the cluster analysis method be listed.</td>
</tr>
</tbody>
</table>
dissimilarity specifies that the dissimilarity measure be listed.
similarity specifies that the similarity measure be listed.
vars specifies that the variables attached to the clusters be listed.
chars specifies that any Stata characteristics attached to the clusters be listed.
other specifies that information attached to the clusters under the heading “other” be listed.

The following option is available with `cluster list` but is not shown in the dialog box:
all, the default, specifies that all items and information attached to the cluster(s) be listed. You may instead pick among the notes, type, method, dissimilarity, similarity, vars, chars, and other options to limit what is presented.

**Options for cluster renamevar**

`name(clname)` indicates the cluster analysis within which the variable renaming is to take place. If `name()` is not specified, the most recently performed cluster analysis (or the one specified by `cluster use`) will be used.

`prefix` specifies that all variables attached to the cluster analysis that have `oldstub` as the beginning of their name be renamed, with `newstub` replacing `oldstub`.

**Remarks and examples**

> Example 1

We demonstrate these `cluster` utility commands by beginning with four already-defined cluster analyses. The `dir` and `list` subcommands provide listings of the cluster analyses.

```
. cluster dir
  bcx3kmed
  ayz5kmeans
  abc_clink
  xyz_slink

. cluster list xyz_slink
  xyz_slink (type: hierarchical, method: single, dissimilarity: L2)
    vars: xyz_slink_id (id variable)
          xyz_slink_ord (order variable)
          xyz_slink_hgt (height variable)
    other: cmd: cluster singlelinkage x y z, name(xyz_slink)
           varlist: x y z
           range: 0 .
```
. cluster list
bcx3kmed (type: partition, method: kmedians, dissimilarity: L2)
  vars: bcx3kmed (group variable)
  other: cmd: cluster kmedians b c x, k(3) name(bcx3kmed)
  varlist: b c x
  k: 3
  start: krandom
  range: 0.
ayz5kmeans (type: partition, method: kmeans, dissimilarity: L2)
  vars: ayz5kmeans (group variable)
  other: cmd: cluster kmeans a y z, k(5) name(ayz5kmeans)
  varlist: a y z
  k: 5
  start: krandom
  range: 0.
abc_clink (type: hierarchical, method: complete, dissimilarity: L2)
  vars: abc_clink_id (id variable)
  abc_clink_ord (order variable)
  abc_clink_hgt (height variable)
  other: cmd: cluster completelinkage a b c, name(abc_clink)
  varlist: a b c
  range: 0.
xyz_slink (type: hierarchical, method: single, dissimilarity: L2)
  vars: xyz_slink_id (id variable)
  xyz_slink_ord (order variable)
  xyz_slink_hgt (height variable)
  other: cmd: cluster singlelinkage x y z, name(xyz_slink)
  varlist: x y z
  range: 0.

. cluster list a*, vars
ayz5kmeans
  vars: ayz5kmeans (group variable)
abc_clink
  vars: abc_clink_id (id variable)
  abc_clink_ord (order variable)
  abc_clink_hgt (height variable)

cluster dir listed the names of the four currently defined cluster analyses. cluster list followed
by the name of one of the cluster analyses listed the information attached to that cluster analysis.
The cluster list command, without an argument, listed the information for all currently defined
cluster analyses. We demonstrated the vars option of cluster list to show that we can restrict
the information that is listed. Notice also the use of a* as the cluster name. The * here indicates that
any ending is allowed. For these four cluster analyses, Stata matches the names ayz5kmeans and
abc_clink.

We now demonstrate the use of the renamevar subcommand.

. cluster renamevar ayz5kmeans g5km
  variable ayz5kmeans not found in bcx3kmed
  r(198);
. cluster renamevar ayz5kmeans g5km, name(ayz5kmeans)
. cluster list ayz5kmeans
ayz5kmeans (type: partition, method: kmeans, dissimilarity: L2)
  vars: g5km (group variable)
  other: cmd: cluster kmeans a y z, k(5) name(ayz5kmeans)
  varlist: a y z
  k: 5
  start: krandom
  range: 0.
The first use of `cluster renamevar` failed because we did not specify which cluster object to use (with the `name()` option), and the most recent cluster object, `bcx3kmed`, was not the appropriate one. After specifying the `name()` option with the appropriate cluster name, the `renamevar` subcommand changed the name as shown in the `cluster list` command that followed.

The cluster `use` command sets a particular cluster object as the default. We show this in conjunction with the `prefix` option of the `renamevar` subcommand.

```plaintext
. cluster use ayz5kmeans
. cluster renamevar g grp, prefix
. cluster renamevar xyz_slink_ wrk, prefix name(xyz*)
. cluster list ayz* xyz*
ayz5kmeans (type: partition, method: kmeans, dissimilarity: L2)
  vars: grp5km (group variable)
    other: cmd: cluster kmeans a y z, k(5) name(ayz5kmeans)
    varlist: a y z
    k: 5
    start: krandom
    range: 0.
xyz_slink (type: hierarchical, method: single, dissimilarity: L2)
  vars: wrkid (id variable)
    wrkord (order variable)
    wrkhgt (height variable)
    other: cmd: cluster singlelinkage x y z, name(xyz_slink)
    varlist: x y z
    range: 0.
```

The cluster `use` command placed `ayz5kmeans` as the current cluster object. The `renamevar` command that followed capitalized on this placement by leaving off the `name()` option. The `prefix` option allowed us to change the variable names, as demonstrated in the `cluster list` of the two changed cluster objects.

`cluster rename` changes the name of cluster objects. `cluster drop` allows us to drop some of or all the cluster objects.

```plaintext
. cluster rename xyz_slink bob
. cluster rename ayz* sam
. cluster list, type method vars
sam (type: partition, method: kmeans)
  vars: grp5km (group variable)
bob (type: hierarchical, method: single)
  vars: wrkid (id variable)
    wrkord (order variable)
    wrkhgt (height variable)
bcx3kmed (type: partition, method: kmedians)
  vars: bcx3kmed (group variable)
abc_clink (type: hierarchical, method: complete)
  vars: abc_clink_id (id variable)
    abc_clink_ord (order variable)
    abc_clink_hgt (height variable)
. cluster drop bcx3kmed abc_clink
. cluster dir
sam
bob
. cluster drop _all
. cluster dir
```
We used options with `cluster list` to limit what was presented. The `_all` keyword with `cluster drop` removed all currently defined cluster objects.

Also see

- [MV] `cluster` — Introduction to cluster-analysis commands
- [MV] `cluster notes` — Cluster analysis notes
- [MV] `cluster programming utilities` — Cluster-analysis programming utilities
- [MV] `clustermat` — Introduction to clustermat commands
- [D] `notes` — Place notes in data
- [P] `char` — Characteristics