

ca postestimation plots — Postestimation plots for ca and camat

[Postestimation commands](#)

[cabiplot](#)

[caprojection](#)

[Remarks and examples](#)

[References](#)

[Also see](#)

Postestimation commands

The following postestimation commands are of special interest after `ca` and `camat`:

Command	Description
cabiplot	biplot of row and column points
caprojection	CA dimension projection plot

cabiplot

Description for cabiplot

`cabiplot` produces a plot of the row points or column points, or a biplot of the row and column points. In this plot, the (Euclidean) distances between row (column) points approximates the χ^2 distances between the associated row (column) profiles if the CA is properly normalized. Similarly, the association between a row and column point is approximated by the inner product of vectors from the origin to the respective points (see [MV] `ca`).

Menu for cabiplot

Statistics > Multivariate analysis > Correspondence analysis > Postestimation after CA > Biplot of row and column points

Syntax for cabiplot

```
cabiplot [ , options ]
```

<i>options</i>	Description
----------------	-------------

Main

<code>dim(# #)</code>	the two dimensions to be displayed; default is <code>dim(2 1)</code>
<code>norow</code>	suppress row coordinates
<code>nocolumn</code>	suppress column coordinates
<code>xnegate</code>	negate the data relative to the <i>x</i> axis
<code>ynegate</code>	negate the data relative to the <i>y</i> axis
<code>maxlength(#)</code>	maximum number of characters for labels; default is <code>maxlength(12)</code>
<code>origin</code>	display the origin on the plot
<code>originlopts(<i>line_options</i>)</code>	affect rendition of origin axes

Rows

<code>rowopts(<i>row_opts</i>)</code>	affect rendition of rows
---------------------------------------	--------------------------

Columns

<code>colopts(<i>col_opts</i>)</code>	affect rendition of columns
---------------------------------------	-----------------------------

Y axis, X axis, Titles, Legend, Overall

<code>twoway_options</code>	any options other than <code>by()</code> documented in [G-3] <code>twoway_options</code>
-----------------------------	--

<i>row_opts</i> and <i>col_opts</i>	Description
-------------------------------------	-------------

<code>plot_options</code>	change look of markers (color, size, etc.) and look or position of marker labels
<code>suppopts(<i>plot_options</i>)</code>	change look of supplementary markers and look or position of supplementary marker labels

<i>plot_options</i>	Description
---------------------	-------------

<code>marker_options</code>	change look of markers (color, size, etc.)
<code>marker_label_options</code>	add marker labels; change look or position

Options for cabiplot

Main

`dim(# #)` identifies the dimensions to be displayed. For instance, `dim(3 2)` plots the third dimension (vertically) versus the second dimension (horizontally). The dimension number cannot exceed the number of extracted dimensions. The default is `dim(2 1)`.

`norow` suppresses plotting of row points.

`nocolumn` suppresses plotting of column points.

`xnegate` specifies that the x -axis values are to be negated (multiplied by -1).

`ynegate` specifies that the y -axis values are to be negated (multiplied by -1).

`maxlength(#)` specifies the maximum number of characters for row and column labels; the default is `maxlength(12)`.

`origin` specifies that the origin be displayed on the plot. This is equivalent to adding the options `xline(0, lcolor(black) lwidth(vthin)) yline(0, lcolor(black) lwidth(vthin))` to the `cabiplot` command.

`originlopts(line_options)` affects the rendition of the origin axes; see [G-3] [line_options](#).

Rows

`rowopts(row_opts)` affects the rendition of the rows. The following *row_opts* are allowed:

plot_options affect the rendition of row markers, including their shape, size, color, and outline (see [G-3] [marker_options](#)) and specify if and how the row markers are to be labeled (see [G-3] [marker_label_options](#)).

`supports(plot_options)` affects supplementary markers and supplementary marker labels; see above for description of *plot_options*.

Columns

`colopts(col_opts)` affects the rendition of columns. The following *col_opts* are allowed:

plot_options affect the rendition of column markers, including their shape, size, color, and outline (see [G-3] [marker_options](#)) and specify if and how the column markers are to be labeled (see [G-3] [marker_label_options](#)).

`supports(plot_options)` affects supplementary markers and supplementary marker labels; see above for description of *plot_options*.

Y axis, X axis, Titles, Legend, Overall

twoway_options are any of the options documented in [G-3] [twoway_options](#), excluding `by()`. These include options for titling the graph (see [G-3] [title_options](#)) and for saving the graph to disk (see [G-3] [saving_option](#)).

`cabiplot` automatically adjusts the aspect ratio on the basis of the range of the data and ensures that the axes are balanced. As an alternative, the *twoway_option* `aspectratio()` can be used to override the default aspect ratio. `cabiplot` accepts the `aspectratio()` option as a suggestion only and will override it when necessary to produce plots with balanced axes; that is, distance on the x axis equals distance on the y axis.

twoway_options, such as `xlabel()`, `xscale()`, `ylabel()`, and `yscale()` should be used with caution. These *axis_options* are accepted but may have unintended side effects on the aspect ratio. See [G-3] [twoway_options](#).

caprojection

Description for caprojection

`caprojection` produces a line plot of the row and column coordinates. The goal of this graph is to show the ordering of row and column categories on each principal dimension of the analysis. Each principal dimension is represented by a vertical line; markers are plotted on the lines where the row and column categories project onto the dimensions.

Menu for caprojection

Statistics > Multivariate analysis > Correspondence analysis > Postestimation after CA > Dimension projection plot

Syntax for caprojection

```
caprojection [ , options ]
```

<i>options</i>	Description
Main	
<code>dim(<i>numlist</i>)</code>	dimensions to be displayed; default is all
<code>norow</code>	suppress row coordinates
<code>nocolumn</code>	suppress column coordinates
<code>alternate</code>	alternate labels
<code>maxlength(#)</code>	number of characters displayed for labels; default is <code>maxlength(12)</code>
<code>combine_options</code>	affect the rendition of the combined column and row graphs
Rows	
<code>rowopts(<i>row_opts</i>)</code>	affect rendition of rows
Columns	
<code>colopts(<i>col_opts</i>)</code>	affect rendition of columns
Y axis, X axis, Titles, Legend, Overall	
<code>twoway_options</code>	any options other than <code>by()</code> documented in [G-3] <code>twoway_options</code>

<i>row_opts</i> and <i>col_opts</i>	Description
<code>plot_options</code>	change look of markers (color, size, etc.) and look or position of marker labels
<code>supportopts(<i>plot_options</i>)</code>	change look of supplementary markers and look or position of supplementary marker labels

<i>plot_options</i>	Description
<code>marker_options</code>	change look of markers (color, size, etc.)
<code>marker_label_options</code>	add marker labels; change look or position

Options for caprojection

Main

`dim(numlist)` identifies the dimensions to be displayed. By default, all dimensions are displayed.

`norow` suppresses plotting of rows.

`nocolumn` suppresses plotting of columns.

`alternate` causes adjacent labels to alternate sides.

`maxlength(#)` specifies the maximum number of characters for row and column labels; the default is `maxlength(12)`.

`combine_options` affect the rendition of the combined plot; see [G-2] [graph combine](#). `combine_options` may not be specified with either `norow` or `nocolumn`.

Rows

`rowopts(row_opts)` affects the rendition of rows. The following *row_opts* are allowed:

`plot_options` affect the rendition of row markers, including their shape, size, color, and outline (see [G-3] [marker_options](#)) and specify if and how the row markers are to be labeled (see [G-3] [marker_label_options](#)).

`suppopts(plot_options)` affects supplementary markers and supplementary marker labels; see above for description of *plot_options*.

Columns

`colopts(col_opts)` affects the rendition of columns. The following *col_opts* are allowed:

`plot_options` affect the rendition of column markers, including their shape, size, color, and outline (see [G-3] [marker_options](#)) and specify if and how the column markers are to be labeled (see [G-3] [marker_label_options](#)).

`suppopts(plot_options)` affects supplementary markers and supplementary marker labels; see above for description of *plot_options*.

Y axis, X axis, Titles, Legend, Overall

`twoway_options` are any of the options documented in [G-3] [twoway_options](#), excluding `by()`. These include options for titling the graph (see [G-3] [title_options](#)) and for saving the graph to disk (see [G-3] [saving_option](#)).

Remarks and examples

[stata.com](http://www.stata.com)

In [example 4](#) of [MV] [ca](#), we showed that plots can be obtained simply by specifying the `plot` option during estimation (or replay). If the default plot is not exactly what you want, the `cabiplot` postestimation command provides control over the appearance of the plot.

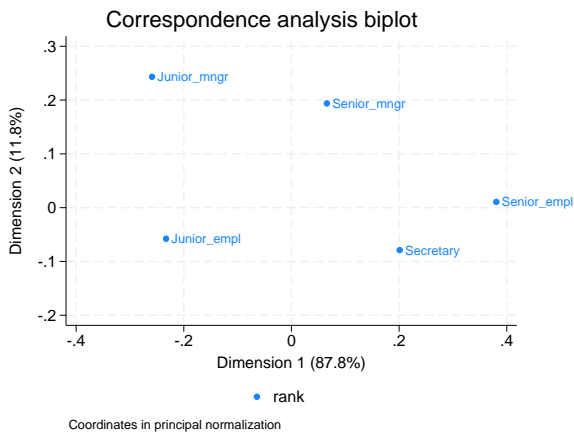
► Example 1: cabiplot

For instance, if we constructed a CA in row principal normalization, we would want to look only at the (points for the) row categories, omitting the column categories. In this normalization, the Euclidean distances between the row points approximate the χ^2 distances between the corresponding row profiles, but the Euclidean distances between the column categories are a distortion of the χ^2 distances of the column profiles. We can use `cabiplot` with the `nocolumn` option to suppress the graphing of the column points.

```

. use https://www.stata-press.com/data/r18/ca_smoking
. quietly ca rank smoking, norm(principal)
. cabiplot, nocolumn legend(on label(1 rank))

```

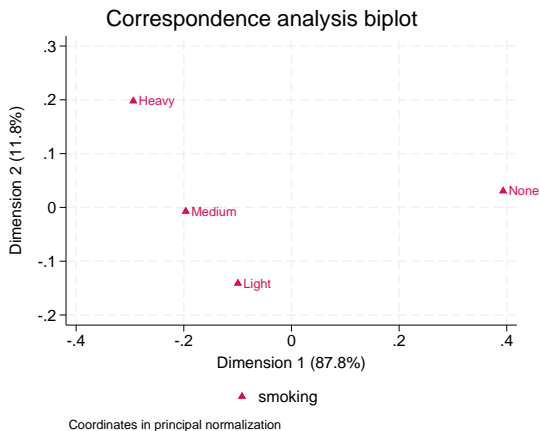


The default graph would not have provided a legend, so we included `legend(on label(1 rank))` to produce one. We see that secretaries have smoking behavior that is rather similar to that of senior employees but rather dissimilar to that of the junior managers, with the other two ranks taking intermediate positions. Because we actually specified the principal normalization, we may also interpret the distances between the smoking categories as approximations to χ^2 distances.

```

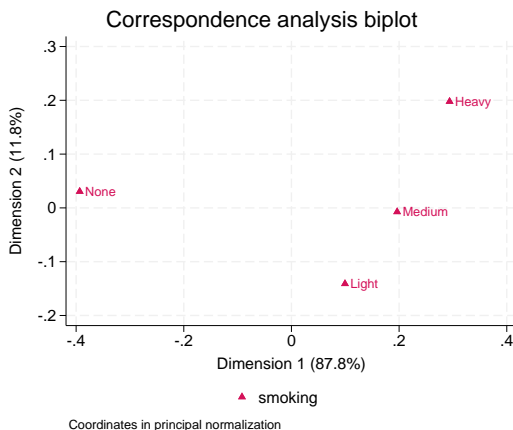
. cabiplot, norow legend(on label(1 smoking))

```



You may not like the orientation of the dimensions. For instance, in this plot, the smokers are on the left and the nonsmokers are on the right. It is more natural to locate the nonsmokers on the left and the smokers on the right so that smoking increases from left to right. This is accomplished with the `xnegate` option.

```
. cabiplot, xnegate norow legend(on label(1 smoking))
```



◀

□ Technical note

To see that negating is possible, think in terms of the fitted values

$$\widehat{P}_{ij} = r_i c_j (1 + \widehat{R}_{i1} \widehat{C}_{j1} + \widehat{R}_{i2} \widehat{C}_{j2} + \dots)$$

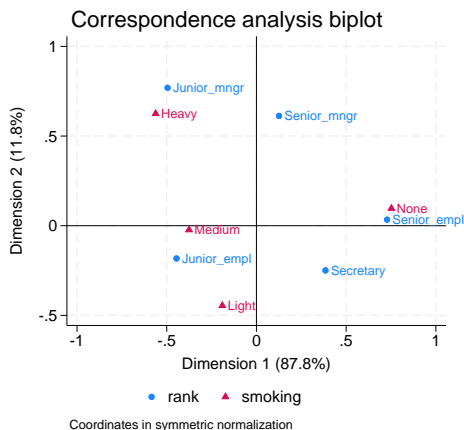
If the sign of the first column of \mathbf{R} and \mathbf{C} is changed at the same time, the fitted values are not affected. This is true for all CA statistics, and it holds true for other columns of \mathbf{R} and \mathbf{C} as well.

□

▷ Example 2: cabiplot with symmetric normalization

Using the symmetric normalization allows us to display a biplot where row categories may be compared with column categories. We execute `ca` again, with the `normalize(symmetric)` option, but suppress the output. This normalization somewhat distorts the interpretation of the distances between row points (or column points) as approximations to χ^2 distances. Thus the similarity of the staff categories (or smoking categories) cannot be adequately assessed. However, this plot allows us to study the association between smoking and rank.

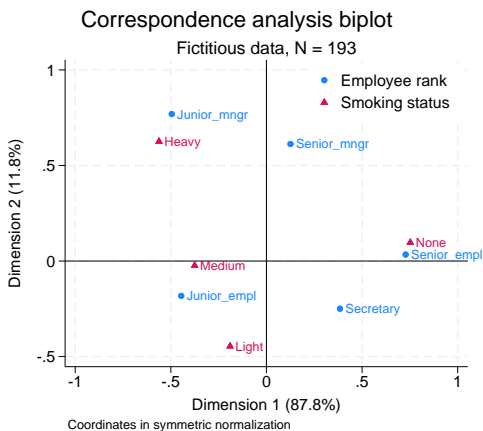
```
. quietly ca rank smoking, normalize(symmetric) dim(2)
. cabiplot, origin
```



With this symmetric normalization, we do not interpret the distances between categories of **smoking** and **rank**. Rather, we have to think in terms of vectors from the origin. The inner product of vectors approximates the residuals from a model of independence of the rows and columns. The inner product depends on the lengths of the vectors and the (cosine of the) angle between the vectors. If the vectors point in the same direction, the residuals are positive—these row and column categories tend to occur together. In our example, we see that senior employees tend to be nonsmokers. If the vectors point in opposite directions, the residuals are negative—these row and column categories tend to be exclusive. In our example, senior managers tend not to be light smokers. Finally, if the vectors are orthogonal (± 90 degrees), the residuals tend to be small; that is, the observed frequencies correspond to what we expect under independence. For instance, junior managers have an average rate of light smoking.

Using various graph options, we can enhance the look of the plot.

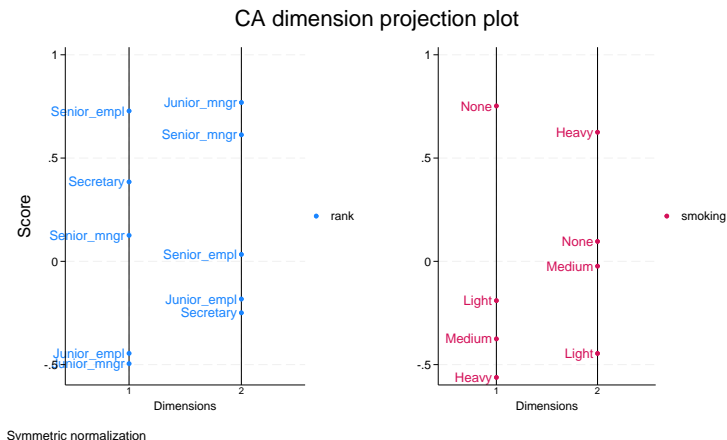
```
. cabiplot, origin subtitle("Fictitious data, N = 193")
> legend(pos(2) ring(0) col(1) lab(1 Employee rank) lab(2 Smoking status))
```



▷ Example 3: caprojection

`caprojection` produces a projection plot of the row and column coordinates after `ca` or `camat` and is especially useful if we think of CA as optimal scaling of the categories of the variables to maximize the correlations between the row and column variables. We continue where we left off with our previous example.

```
. caprojection
```



This example has relatively few categories, so we could visualize the orderings of the rows and columns from the previous biplots. However, CA is often used with larger problems, and in those cases, a projection plot is a useful presentation device.

◀

References

See [References](#) in `[MV] ca`.

Also see

`[MV] ca` — Simple correspondence analysis

`[MV] ca postestimation` — Postestimation tools for `ca` and `camat`

`[MV] screeplot` — Scree plot of eigenvalues

Stata, Stata Press, and Mata are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2023 StataCorp LLC, College Station, TX, USA. All rights reserved.

