

Title

intro — Introduction to multivariate statistics manual

Description

This entry describes this manual and what has changed since Stata 9.

Remarks

This manual documents Stata's multivariate analysis features and is referred to as the [MV] manual in cross-references.

Following this entry, [MV] **multivariate** provides an overview of the multivariate analysis features in Stata and Stata's multivariate analysis commands. The other parts of this manual are arranged alphabetically.

Stata is continually being updated, and Stata users are always writing new commands. To find out about the latest multivariate analysis features, type `search multivariate analysis` after installing the latest official updates; see [R] **update**.

What's new

This section is intended for previous Stata users. If you are new to Stata, you may as well skip it.

1. New estimation commands `discrim` and `candisc` provide several discriminant analysis techniques, including linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), logistic discriminant analysis, and k th-nearest-neighbor (KNN) discrimination. See [MV] **discrim**, [MV] **discrim estat**, and [MV] **candisc**.
2. Existing estimation commands `mds`, `mdslong`, and `mdsmat` now provide modern as well as classical multidimensional scaling (MDS), including metric and nonmetric MDS. Available loss functions include stress, normalized stress, squared stress, normalized squared stress, and Sammon. Available transformations include identity, power, and monotonic. `mdslong` also now allows `aweights` and `fweights`, and `mdsmat` has a `weight()` option. See [MV] **mds**, [MV] **mdslong**, and [MV] **mdsmat**.
3. New estimation command `mca` provides multiple correspondence analysis (MCA) and joint correspondence analysis (JCA); see [MV] **mca** and [MV] **mca postestimation**. You can use existing command `screeplot` afterward to graph principal inertias; see [MV] **screeplot**.
4. Concerning existing estimation command `ca` (correspondence analysis),
 - a. `ca` now allows crossed (stacked) variables. This provides a way to automatically combine two or more categorical variables into one crossed variable and perform correspondence analysis with it.
 - b. `ca`'s existing option `normalize()` now allows `normalize(standard)` to provide normalization of the coordinates by singular vectors divided by the square root of the mass.
 - c. `ca`'s new option `length()` allows you to customize the length of labels with crossed variables in output.
 - d. New postestimation command `estat loadings`, used after `ca` and `camat`, displays correlations of profiles and axes.

- e. Existing postestimation command `cabiplot` has new option `origin` that displays the origin within the plot. `cabiplot` also now accepts `originlopts` (*line_options*) to customize the appearance of the origin on the graph.
- f. Existing postestimation commands `cabiplot` and `caprojection` now allow row and column marker labels to be specified using the `mlabel()` suboption of `rowopts()` and `colopts()`.

See [MV] **ca** and [MV] **ca postestimation**.

5. Existing commands `cluster`, `matrix dissimilarity`, and `mds` now allow the Gower measure for a mix of binary and continuous data; see [MV] *measure_option*.
6. Existing command `biplot` has new options. `dim()` specifies the dimensions to be displayed. `negcol` specifies that negative column (variable) arrows be plotted. `negcolopts` (*col_options*) provides graph options for the negative column arrows. `norow` and `nocolumn` suppress the row points or column arrows. See [MV] **biplot**.
7. New postestimation command `estat rotate` after `canon` performs orthogonal varimax rotation of the raw coefficients, standard coefficients, or canonical loadings. After `estat rotate`, new postestimation command `estat rotatecompare` displays the rotated and unrotated coefficients or loadings and the most recently rotated coefficients or loadings. See [MV] **canon postestimation**.
8. Existing commands `pcamat` and `factormat` now allow singular correlation or covariance structures. New option `forcepsd` modifies a matrix to be positive semidefinite and thus to be a proper covariance matrix. See [MV] **pca** and [MV] **factor**.
9. Existing commands `rotate` and `rotatemat` now refer to “Kaiser normalization” rather than “Horst normalization”. A search of the literature indicates that Kaiser normalization is the preferred terminology. Previously option `horst` was a synonym for `normalize`. Now option `horst` is not documented. See [MV] **rotate** and [MV] **rotatemat**.
10. Existing command `procrustes` now saves the number of y variables in scalar `e(ny)`; see [MV] **procrustes**.

For a complete list of all the new features in Stata 10, see [U] **1.3 What’s new**.

Also See

[U] **1.3 What’s new**

[R] **intro** — Introduction to base reference manual