Title

intro — Introduction to multivariate statistics manual

Description

This entry describes this manual and what has changed since Stata 10.

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 command `mvtest` performs multivariate tests on means, covariances, and correlations (both one-sample and multiple-sample), and it performs tests of univariate, bivariate, and multivariate normality. Included are Box's M test for covariances, and for tests of normality, the Doornik–Hansen omnibus test, Henze–Zirkler test, Mardia's multivariate kurtosis test, and Mardia's multivariate skewness test. See [MV] **mvtest**.
2. The new factor-variable syntax allowed throughout Stata affects `manova` even though `manova` always allowed factor variables. See [MV] **manova**.
   a. `manova` has an all-new syntax. The old syntax continues to work under version control.
   b. `manova`, just like `anova`, adopts the new factor-variable syntax, but with a twist. In other Stata commands, continuous is assumed and you use `i.`*varname* to indicate a categorical variable. In `manova` and `anova`, categorical is assumed and you use `c.`*varname* to indicate continuous. Thus the options `category()`, `class()`, and `continuous()` are no longer used.
   c. To form an interaction, you use *varname1*`#`*varname2*. Previously, you used *varname1*`*`*varname2*. `*` now means variable-name expansion, just as it does on other commands, so you could type `manova y* = a b* a#b*`. The `|` symbol continues to be used for nesting.
   d. You can now use *varname1*`##`*varname2* as a shorthand for full factorial, meaning *varname1* *varname2* *varname1*`#`*varname2*. You can use *varname1*`##`*varname2*`##`*varname3* for 3-way factorial, and so on.
3. Existing command `mvreg` may now be used after `manova` to show results in regression-style format, just as `regress` can be used after `anova`. See [MV] **manova**.
4. Existing command `test` after `manova`, in addition to allowing the special syntax previously provided, now allows all the standard `test` syntax, too. See [MV] **manova postestimation**.

5. Existing commands `predictnl`, `nlcom`, `testnl`, and `testparm` may now be used after `manova`; see [R] **predictnl**, [R] **nlcom**, [R] **testnl**, and [R] **test**.
6. New postestimation command `margins` may be used after `manova`. See [R] **margins**.
7. `manova` now requires that categorical variables take on nonnegative integer values. Previously, a categorical variable could take on values −1, 2.5, 3.14159, etc., although few did. Arbitrary values are still allowed under version control. See [MV] **manova**.
8. `manova`'s new option `dropemptycells` removes unobserved levels from the model rather than setting their coefficients to zero. Statistically, the approaches are equivalent. Computationally, a larger `matsize` is required when empty cells are retained. In models with many interactions, you may need to specify this option. See [MV] **manova** and see [R] **set emptycells**.
9. Programmers: The row and column names on `e(b)`, `e(V)`, etc., after `manova` are now meaningful and follow standard factor-variable notation. See *What's new* in [P] **intro**.
10. Existing command `biplot` has several improvements:
    a. `biplot` can now be used with larger datasets. Previously, the row dimension was limited by Stata's maximum `matsize`.
    b. `biplot` has new option `generate()`, which saves the coordinates of observations in variables.
    c. `biplot` has new options `rowover()` and `row#opts()`, which allow highlighting groups of observations on the graph and customizing the look of the graph.
    d. New option `rowlabel()` makes customizing rows easier.
    e. `biplot` now drops constant variables from the computation.
    f. `biplot` now uses an improved version of the singular value decomposition, which may result in sign differences and slight differences in values.
    g. `rowopts()`, `colopts()`, and `negcolopts()` now allow names to contain simple and compound quotes.
    h. `biplot` did not honor option `scheme(economist)` for separate graphs (option `separate`). This has been fixed.
11. Existing command `canon`'s default output has changed. It previously displayed something that looked like estimation output but was not because standard errors were conditional. The output now looks like you would expect. The conditional output can be obtained by specifying new option `stderr` or under version control (`set version` to 10 or before).
12. The manual now includes a glossary; see [MV] **Glossary**.

For a complete list of all the new features in Stata 11, see [U] **1.3 What's new**.

Also see

[U] **1.3 What's new**

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