Title [stata.com]

> **mds postestimation** — Postestimation tools for mds, mdsmat, and mdslong

Postestimation commands

The following postestimation commands are of special interest after mds, mdslong, and mdsmat:

Command	Description
estat config	coordinates of the approximating configuration
estat correlations	correlations between dissimilarities and approximating distances
estat pairwise	pairwise dissimilarities, approximating distances, and raw residuals
estat quantiles	quantiles of the residuals per object
estat stress	Kruskal stress (loss) measure (only after classical MDS)
* estat summarize	estimation sample summary
mdsconfig	plot of approximating configuration
mdsshepard	Shepard diagram
screeplot	plot eigenvalues (only after classical MDS)

*estat summarize is not available after mdsmat.

The following standard postestimation commands are also available:

Command	Description
* estimates	cataloging estimation results
predict	approximating configuration and pairwise statistics

*All estimates subcommands except table and stats are available.

predict

Description for predict

predict creates new variables containing predictions such as approximating configurations in Euclidean space and selected pairwise statistics.

Menu for predict

Statistics > Postestimation

Syntax for predict

    predict [*type*] {*stub*∗ | *newvarlist*} [*if*] [*in*] [, *statistic options*]

statistic	Description
<u>Main</u>	
<u>config</u>	approximating configuration; specify dimension() or fewer variables
<u>pairwise</u>(*pstats*)	selected pairwise statistics; specify same number of variables

pstats	Description
<u>dispar</u>ities	disparities = transformed(dissimilarities)
<u>diss</u>imilarities	dissimilarities
<u>dist</u>ances	Euclidean distances between configuration points
<u>rr</u>esiduals	raw residual = dissimilarity − distance
<u>tr</u>esiduals	transformed residual = disparity − distance
<u>weights</u>	weights

options	Description
<u>Main</u>	
* <u>saving</u>(*filename* , replace)	save results to *filename*; use replace to overwrite existing *filename*
full	create predictions for all pairs of objects; pairwise() only

∗ saving() is required after mdsmat, after mds if pairwise() is selected, and after mdslong if config is selected.

Options for predict

    ┌─ Main ───

config generates variables containing the approximating configuration in Euclidean space. Specify as many new variables as approximating dimensions (as determined by the dimension() option of mds, mdsmat, or mdslong), though you may specify fewer. estat config displays the same information but does not store the information in variables. After mdsmat and mdslong, you must also specify the saving() option.

pairwise(*pstats*) generates new variables containing pairwise statistics. The number of new variables should be the same as the number of specified statistics. The following statistics are allowed:

disparities generates the disparities, that is, the transformed dissimilarities. If no transformation is applied (modern MDS with transform(identity)), disparities are the same as dissimilarities.

dissimilarities generates the dissimilarities used in MDS. If mds, mdslong, or mdsmat was invoked on similarity data, the associated dissimilarities are returned.

distances generates the (unsquared) Euclidean distances between the fitted configuration points.

rresiduals generates the raw residuals: dissimilarities − distances.

tresiduals generates the transformed residuals: disparities − distances.

weights generates the weights. Missing proximities are represented by zero weights.

estat pairwise displays some of the same information but does not store the information in variables.

After mds and mdsmat, you must also specify the saving() option. With n objects, the pairwise dataset has $n(n-1)/2$ observations. In addition to the three requested variables, predict produces variables *id*1 and *id*2, which identify pairs of objects. With mds, *id* is the name of the identification variable (id() option), and with mdsmat, it is "Category".

saving(*filename*[, replace]) is required after mdsmat, after mds if pairwise() is selected, and after mdslong if config is selected. saving() indicates that the generated variables are to be created in a new Stata dataset and saved in the file named *filename*. Unless saving() is specified, the variables are generated in the current dataset.

replace indicates that *filename* specified in saving() may be overwritten.

full creates predictions for all pairs of objects (j_1, j_2). The default is to generate predictions only for pairs (j_1, j_2) where $j_1 > j_2$. full may be specified only with pairwise().

estat

Description for estat

estat config lists the coordinates of the approximating configuration.

estat correlations lists the Pearson and Spearman correlations between the disparities or dissimilarities and the Euclidean distances for each object.

estat pairwise lists the pairwise statistics: the disparities, the distances, and the residuals.

estat quantiles lists the quantiles of the residuals per object.

estat stress displays the Kruskal stress (loss) measure between the (transformed) dissimilarities and fitted distances per object (only after classical MDS).

estat summarize summarizes the variables in the MDS over the estimation sample. After mds, estat summarize also reports whether and how variables were transformed before computing similarities or dissimilarities.

Menu for estat

Statistics > Postestimation

Syntax for estat

List the coordinates of the approximating configuration

    estat <u>con</u>fig [, <u>max</u>length(*#*) <u>for</u>mat(%*fmt*)]

List the Pearson and Spearman correlations

    estat <u>cor</u>relations [, <u>max</u>length(*#*) <u>for</u>mat(%*fmt*) <u>notrans</u>form <u>notot</u>al]

List the pairwise statistics: disparities, distances, and residuals

    estat <u>pair</u>wise [, <u>max</u>length(*#*) <u>notrans</u>form <u>full</u> <u>separa</u>tor]

List the quantiles of the residuals

    estat <u>qua</u>ntiles [, <u>max</u>length(*#*) <u>for</u>mat(%*fmt*) <u>notot</u>al <u>notrans</u>form]

Display the Kruskal stress (loss) measure per point (only after classical MDS)

    estat <u>str</u>ess [, <u>max</u>length(*#*) <u>for</u>mat(%*fmt*) <u>notot</u>al <u>notrans</u>form]

Summarize the variables in MDS

    estat <u>sum</u>marize [, <u>labels</u>]

options	Description
<u>max</u>length(*#*)	maximum number of characters for displaying object names; default is 12
<u>for</u>mat(%*fmt*)	display format
<u>notot</u>al	suppress display of overall summary statistics
<u>notrans</u>form	use dissimilarities instead of disparities
<u>full</u>	display all pairs (j_1, j_2); default is $(j_1 > j_2)$ only
<u>separa</u>tor	draw separating lines
<u>labels</u>	display variable labels

collect is allowed with estat correlations, estat quantiles, estat stress, and estat summarize; see
    [U] **11.1.10 Prefix commands**.

Options for estat

maxlength(*#*), an option used with all but estat summarize, specifies the maximum number of
    characters of the object names to be displayed; the default is maxlength(12).

format(%*fmt*), an option used with estat config, estat correlations, estat quantiles, and
    estat stress, specifies the display format; the default differs between the subcommands.

nototal, an option used with estat correlations, estat quantiles, and estat stress, suppresses the overall summary statistics.

notransform, an option used with estat correlations, estat pairwise, estat quantiles, and estat stress, specifies that the untransformed dissimilarities be used instead of the transformed dissimilarities (disparities).

full, an option used with estat pairwise, displays a row for all pairs (j_1, j_2). The default is to display rows only for pairs where $j_1 > j_2$.

separator, an option used with estat pairwise, draws separating lines between blocks of rows corresponding to changes in the first of the pair of objects.

labels, an option used with estat summarize, displays variable labels.

Remarks and examples

stata.com

Remarks are presented under the following headings:

>*Postestimation statistics*
>*Predictions*

Postestimation statistics

After an MDS analysis, several facilities can help you better understand the analysis and, in particular, to assess the quality of the lower-dimensional Euclidean representation. We display results after classical MDS. All are available after modern MDS except for estat stress.

▷ Example 1: estat

We illustrate the MDS postestimation facilities with the Morse code digit-similarity dataset; see example 1 in [MV] **mdslong**.

```
. use https://www.stata-press.com/data/r18/morse_long
(Morse data (Rothkopf 1957))
. generate sim = freqsame/100
. mdslong sim, id(digit1 digit2) s2d(standard) noplot
 (output omitted)
```

MDS has produced a two-dimensional configuration with Euclidean distances approximating the dissimilarities between the Morse codes for digits. This configuration may be examined using the estat config command; see mdsconfig if you want to plot the configuration.

```
. estat config
Configuration in 2-dimensional Euclidean space (principal normalization)
```

digit1	dim1	dim2
0	0.5690	0.0162
1	0.4561	-0.3384
2	0.0372	-0.5854
3	-0.3878	-0.4516
4	-0.5800	-0.0770
5	-0.5458	-0.0196
6	-0.3960	0.4187
7	-0.0963	0.5901
8	0.3124	0.3862
9	0.6312	0.0608

This configuration is not unique. A translation, a reflection, and an orthonormal rotation of the configuration do not affect the interpoint Euclidean distances. All such transformations are equally reasonable MDS solutions. Thus you should not interpret aspects of these numbers (or of the configuration plot) that are not invariant to these transformations.

We now turn to the three `estat` subcommands that analyze the MDS residuals, that is, the differences between the disparities or dissimilarities and the matching Euclidean distances. There is a catch here. The raw residuals of MDS are not well behaved. For instance, the sum of the raw residuals is not zero—often it is not even close. The MDS solution does *not* minimize the sum of squares of the raw residuals Mardia, Kent, and Bibby (1979, 406–408). To create reasonable residuals with MDS, the dissimilarities can be transformed to disparities approximating the Euclidean distances. In classical MDS we use a linear transform f, fit by least squares. This is equivalent to Kruskal's Stress1 loss function. The modified residuals are defined as the differences between the linearly transformed dissimilarities and the matching Euclidean distances.

In modern MDS we have three types of transformations from dissimilarities to disparities to choose from: the identity (which does not transform the dissimilarities), a power transformation, and a monotonic transformation.

The three `estat` subcommands summarize the residuals in different ways. After classical MDS, `estat stress` displays the Kruskal loss or stress measures for each object and the overall total.

```
. estat stress
```
Stress between disparities and Euclidean distances

digit1	Kruskal
0	0.1339
1	0.1255
2	0.1972
3	0.2028
4	0.2040
5	0.2733
6	0.1926
7	0.1921
8	0.1715
9	0.1049
Total	0.1848

Second, after classical or modern MDS, the quantiles of the residuals are available, both overall and for the subgroup of object pairs in which an object is involved.

```
. estat quantiles
```
Quantiles of transformed residuals

digit1	N	min	p25	q50	q75	max
0	9	-.111732	-.088079	-.028917	.11202	.220399
1	9	-.170063	-.137246	-.041244	.000571	.11202
2	9	-.332717	-.159472	-.072359	.074999	.234867
3	9	-.136251	-.120398	-.072359	.105572	.365833
4	9	-.160797	-.014099	.03845	.208215	.355053
5	9	-.09971	-.035357	.176337	.325043	.365833
6	9	-.137246	-.113564	-.075008	.177448	.325043
7	9	-.332717	-.170063	-.124129	.03845	.176337
8	9	-.186452	-.134831	-.041244	.075766	.220399
9	9	-.160797	-.104403	-.088079	-.064316	-.030032
Total	90	-.332717	-.113564	-.041244	.105572	.365833

The dissimilarities for the Morse code of digit 5 are fit considerably worse than for all other Morse codes. Digit 5 has the largest Kruskal stress (0.273) and median residual (0.176).

Finally, after classical or modern MDS, estat correlations displays the Pearson and Spearman correlations between the (transformed or untransformed) dissimilarities and the Euclidean distances.

```
. estat correlations
```

Correlations of disparities and Euclidean distances

digit1	N	Pearson	Spearman
0	9	0.9510	0.9540
1	9	0.9397	0.7782
2	9	0.7674	0.4017
3	9	0.7922	0.7815
4	9	0.9899	0.9289
5	9	0.9412	0.9121
6	9	0.8226	0.8667
7	9	0.8444	0.4268
8	9	0.8505	0.7000
9	9	0.9954	0.9333
Total	90	0.8602	0.8301

◁

Predictions

It is possible to generate variables containing the results from the MDS analysis. MDS operates at two levels: first at the level of the objects and second at the level of relations between the objects or pairs of objects. You can generate variables at both of these levels.

The config option of predict after an MDS requests that the coordinates of the objects in the matching configuration be stored in variables. You may specify as many variables as there are retained dimensions. You may also specify fewer variables. The first variable will store the coordinates from the first dimension. The second variable, if any, will store the coordinates from the second dimension, and so on.

The pairwise() option specifies that a given selection of the pairwise statistics are stored in variables. The statistics available are the disparities, dissimilarities, fitted distances, raw residuals, transformed residuals, and weights. The raw residuals are the difference between dissimilarities and the fitted distances, and the transformed residuals are the difference between the disparities and the fitted distances.

There is a complicating issue. With n objects, there are $n(n-1)/2$ pairs of objects. So, to store properties of objects, you need n observations, but to store properties of pairs of objects, you need $n(n-1)/2$ observations. So, where do you store the variables? predict after MDS commands can save the predicted variables in a new dataset. Specify the option saving(*filename*). Such a dataset will automatically have the appropriate object identification variable or variables.

Sometimes it is also possible to store the variables in the dataset you have in memory: object-level variables in an object-level dataset and pairwise-level variables in a pairwise-level dataset.

After mds you have a dataset in memory in which the observations are associated with the MDS objects. Here you can store object-level variables directly in the dataset in memory. To do so, you just omit the saving() option. Here it is not possible to store the pairwise statistics in the dataset in memory. The pairwise statistics have to be stored in a new dataset.

After `mdslong`, the dataset in memory is in a pairwise form, so the variables predicted with the option `pairwise()` can be stored in the dataset in memory. It is, of course, also possible to store the pairwise variables in a new dataset; the choice is yours. With pairwise data in memory, you cannot store the object-level predicted variables into the data in memory; you need to specify the name of a new dataset.

After `mdsmat`, you always need to save the predicted variables in a new dataset.

▷ Example 2: Predictions

Continuing with example 1; the dataset in memory is in long form. Thus we can store the pairwise statistics with the dataset in memory.

```
. predict tdissim eudist resid, pairwise
. list in 1/10
```

	digit1	digit2	freqsame	sim	tdissim	eudist	resid
1.	2	1	62	.62	.3227682	.4862905	-.1635224
2.	3	1	16	.16	.957076	.851504	.105572
3.	3	2	59	.59	.3732277	.4455871	-.0723594
4.	4	1	6	.06	1.069154	1.068583	.0005709
5.	4	2	23	.23	.8745967	.7995979	.0749988
6.	4	3	38	.38	.6841489	.4209922	.2631567
7.	5	1	12	.12	1.002667	1.051398	-.048731
8.	5	2	8	.08	1.047234	.8123672	.2348665
9.	5	3	27	.27	.8257753	.4599419	.3658335
10.	5	4	56	.56	.4218725	.0668193	.3550532

Because we used `mdslong`, the object-level statistics must be saved in a file.

```
. predict d1 d2, config saving(digitdata)
. describe digit* d1 d2 using digitdata
```

Variable name	Storage type	Display format	Value label	Variable label
digit1	byte	%9.0g		
d1	float	%9.0g		MDS dimension 1
d2	float	%9.0g		MDS dimension 2

The information in these variables was already shown with `estat config`. The dataset created has variables `d1` and `d2` with the coordinates of the Morse digits on the two retained dimensions and an identification variable `digit1`. Use `merge` to add these variables to the data in memory; see [D] **merge**.

◁

Stored results

`estat correlations` stores the following in `r()`:

Matrices
    r(R)   statistics per object; columns with # of obs., Pearson corr., and Spearman corr.
    r(T)   overall statistics; # of obs., Pearson corr., and Spearman corr.

`estat quantiles` stores the following in `r()`:

Macros
    r(dtype)   adjusted or raw; dissimilarity transformation
Matrices
    r(Q)   statistics per object; columns with # of obs., min., p25, p50, p75, and max.
    r(T)   overall statistics; # of obs., min., p25, p50, p75, and max.

`estat stress` stores the following in `r()`:

Macros
    r(dtype)   adjusted or raw; dissimilarity transformation
Matrices
    r(S)   Kruskal's stress/loss measure per object
    r(T) 1×1 matrix with the overall Kruskal stress/loss measure

Methods and formulas

See [MV] **mdsmat** for information on the methods and formulas for multidimensional scaling.

For classical MDS, let D_{ij} be the dissimilarity between objects i and j, $1 \le i, j \le n$. We assume $D_{ii} = 0$ and $D_{ij} = D_{ji}$. Let E_{ij} be the Euclidean distance between rows i and j of the matching configuration $\mathbf{Y}$. In classical MDS, $\mathbf{D} - \mathbf{E}$ is not a well-behaved residual matrix. We follow the approach used in metric and nonmetric MDS to transform D_{ij} to "optimally match" E_{ij}, with $\widehat{D}_{ij} = a + bD_{ij}$, where a and b are chosen to minimize the residual sum of squares. This is a simple regression problem and is equivalent to minimizing Kruskal's stress measure (Kruskal 1964; Cox and Cox 2001, 63)

$$\text{Kruskal}(\widehat{\mathbf{D}}, \mathbf{E}) = \left\{ \frac{\sum (E_{ij} - \widehat{D}_{ij})^2}{\sum E_{ij}^2} \right\}^{1/2}$$

with summation over all pairs (i, j). We call the $\widehat{D}_{ij}$ the adjusted or transformed dissimilarities. If the transformation step is skipped by specifying the option `notransform`, we set $\widehat{D}_{ij} = D_{ij}$.

In `estat stress`, the decomposition of Kruskal's stress measure over the objects is displayed. $\text{Kruskal}(\widehat{\mathbf{D}}, \mathbf{E})_i$ is defined analogously with summation over all $j \neq i$.

For modern MDS, the optimal transformation to disparities, $f(\mathbf{D}) \to \widehat{\mathbf{D}}$, is calculated during the estimation. See [MV] **mdsmat** for details. For `transform(power)`, the power is stored in `e(alpha)`. For `transform(monotonic)`, the disparities themselves are stored as `e(Disparities)`.

References

Borg, I., and P. J. F. Groenen. 2005. *Modern Multidimensional Scaling: Theory and Applications.* 2nd ed. New York: Springer.

Cox, T. F., and M. A. A. Cox. 2001. *Multidimensional Scaling.* 2nd ed. Boca Raton, FL: Chapman and Hall/CRC.

Kruskal, J. B. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29: 1–27. https://doi.org/10.1007/bf02289565.

Mardia, K. V., J. T. Kent, and J. M. Bibby. 1979. *Multivariate Analysis.* London: Academic Press.

Also see *References* in [MV] **mdsmat**.

Also see

[MV] **mds** — Multidimensional scaling for two-way data

[MV] **mdslong** — Multidimensional scaling of proximity data in long format

[MV] **mdsmat** — Multidimensional scaling of proximity data in a matrix

[MV] **mds postestimation plots** — Postestimation plots for mds, mdsmat, and mdslong

[MV] **screeplot** — Scree plot of eigenvalues

[U] **20 Estimation and postestimation commands**