Title

stata.com

discrim qda postestimation — Postestimation tools for discrim qda

Postestimation commands

The following postestimation commands are of special interest after `discrim qda`:

Command	Description
`estat classtable`	classification table
`estat correlations`	group correlation matrices and p-values
`estat covariance`	group covariance matrices
`estat errorrate`	classification error-rate estimation
`estat grdistances`	Mahalanobis and generalized squared distances between the group means
`estat grsummarize`	group summaries
`estat list`	classification listing
`estat summarize`	estimation sample summary

The following standard postestimation commands are also available:

Command	Description
* `estimates`	cataloging estimation results
`predict`	group classification and posterior probabilities

* All `estimates` subcommands except `table` and `stats` are available; see [R] **estimates**.

predict

Description for predict

`predict` creates a new variable containing predictions such as group classifications, probabilities, Mahalanobis squared distances, leave-one-out group classifications, leave-one-out probabilities, and leave-one-out Mahalanobis squared distances.

Menu for predict

Statistics > Postestimation

Syntax for predict

`predict` [*type*] *newvar* [*if*] [*in*] [`,` *statistic options*]

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

statistic	Description
Main	
`classification`	group membership classification; the default when one variable is specified and `group()` is not specified
`pr`	probability of group membership; the default when `group()` is specified or when multiple variables are specified
`mahalanobis`	Mahalanobis squared distance between observations and groups
`clscore`	group classification function score
* `looclass`	leave-one-out group membership classification; may be used only when one new variable is specified
* `loopr`	leave-one-out probability of group membership
* `loomahal`	leave-one-out Mahalanobis squared distance between observations and groups

options	Description
Main	
`group(`*group*`)`	the group for which the statistic is to be calculated
Options	
`priors(`*priors*`)`	group prior probabilities; defaults to `e(grouppriors)`
`ties(`*ties*`)`	how ties in classification are to be handled; defaults to `e(ties)`

priors	Description
`equal`	equal prior probabilities
`proportional`	group-size-proportional prior probabilities
matname	row or column vector containing the group prior probabilities
matrix_exp	matrix expression providing a row or column vector of the group prior probabilities

ties	Description
`missing`	ties in group classification produce missing values
`random`	ties in group classification are broken randomly
`first`	ties in group classification are set to the first tied group

You specify one new variable with `classification` or `looclass` and specify either one or `e(N_groups)` new variables with `pr`, `loopr`, `mahalanobis`, `loomahal`, or `clscore`.

Unstarred statistics are available both in and out of sample; type `predict ... if e(sample) ...` if wanted only for the estimation sample. Starred statistics are calculated only for the estimation sample, even when `if e(sample)` is not specified.

`group()` is not allowed with `classification` or `looclass`.

Options for predict

Main

`classification`, the default, calculates the group classification. Only one new variable may be specified.

`pr` calculates group membership posterior probabilities. If you specify the `group()` option, specify one new variable. Otherwise, you must specify `e(N_groups)` new variables.

`mahalanobis` calculates the squared Mahalanobis distance between the observations and group means. If you specify the `group()` option, specify one new variable. Otherwise, you must specify `e(N_groups)` new variables.

`clscore` produces the group classification function score. If you specify the `group()` option, specify one new variable. Otherwise, you must specify `e(N_groups)` new variables.

`looclass` calculates the leave-one-out group classifications. Only one new variable may be specified. Leave-one-out calculations are restricted to `e(sample)` observations.

`loopr` calculates the leave-one-out group membership posterior probabilities. If you specify the `group()` option, specify one new variable. Otherwise, you must specify `e(N_groups)` new variables. Leave-one-out calculations are restricted to `e(sample)` observations.

`loomahal` calculates the leave-one-out squared Mahalanobis distance between the observations and group means. If you specify the `group()` option, specify one new variable. Otherwise, you must specify `e(N_groups)` new variables. Leave-one-out calculations are restricted to `e(sample)` observations.

`group(`*group*`)` specifies the group for which the statistic is to be calculated and can be specified using

`#1`, `#2`, ..., where `#1` means the first category of the `e(groupvar)` variable, `#2` the second category, etc.;

the values of the `e(groupvar)` variable; or

the value labels of the `e(groupvar)` variable if they exist.

`group()` is not allowed with `classification` or `looclass`.

Options

`priors(`*priors*`)` specifies the prior probabilities for group membership. If `priors()` is not specified, `e(grouppriors)` is used. The following *priors* are allowed:

`priors(equal)` specifies equal prior probabilities.

`priors(proportional)` specifies group-size-proportional prior probabilities.

`priors(`*matname*`)` specifies a row or column vector containing the group prior probabilities.

`priors(`*matrix_exp*`)` specifies a matrix expression providing a row or column vector of the group prior probabilities.

`ties(`*ties*`)` specifies how ties in group classification will be handled. If `ties()` is not specified, `e(ties)` is used. The following *ties* are allowed:

`ties(missing)` specifies that ties in group classification produce missing values.

`ties(random)` specifies that ties in group classification are broken randomly.

`ties(first)` specifies that ties in group classification are set to the first tied group.

estat

Description for estat

`estat correlations` displays group correlation matrices. Two-tailed p-values for the correlations are also available.

`estat covariance` displays group covariance matrices.

`estat grdistances` provides Mahalanobis squared distances and generalized squared distances between the group means.

Menu for estat

Statistics > Postestimation

Syntax for estat

Group correlation matrices and p-values

`estat correlations` [, *correlations_options*]

Group covariance matrices

`estat covariance` [, *covariance_options*]

Mahalanobis and generalized squared distances between the group means

`estat grdistances` [, *grdistances_options*]

correlations_options	Description
Main	
`p`	display two-sided p-values
`format(%`*fmt*`)`	numeric display format; default is `%9.0g`
`nohalf`	display full matrix even if symmetric

covariance_options	Description
Main	
`format(%`*fmt*`)`	numeric display format; default is `%9.0g`
`nohalf`	display full matrix even if symmetric

grdistances_options	Description
Main	
mahalanobis	display Mahalanobis squared distances between group means; the default
generalized	display generalized Mahalanobis squared distances between group means
all	equivalent to mahalanobis generalized
format(%*fmt*)	numeric display format; default is %9.0g
Options	
priors(*priors*)	group prior probabilities; defaults to e(grouppriors)

Options for estat

Options for estat are presented under the following headings:

Options for estat correlations
Options for estat covariance
Options for estat grdistances

Options for estat correlations

Main

p specifies that two-sided p-values be computed and displayed for the correlations.

format(%*fmt*) specifies the matrix display format. The default is format(%8.5f).

nohalf specifies that, even though the matrix is symmetric, the full matrix be printed. The default is to print only the lower triangle.

Options for estat covariance

Main

format(%*fmt*) specifies the matrix display format. The default is format(%9.0g).

nohalf specifies that, even though the matrix is symmetric, the full matrix be printed. The default is to print only the lower triangle.

Options for estat grdistances

Main

mahalanobis specifies that a table of Mahalanobis squared distances between group means be presented.

generalized specifies that a table of generalized Mahalanobis squared distances between group means be presented. generalized starts with what is produced by the mahalanobis option and adds a term for the possibly unequal covariances and a term accounting for prior probabilities. Prior probabilities are provided with the priors() option, or if priors() is not specified, by the values in e(grouppriors). By common convention, if prior probabilities are equal across the groups, the prior probability term is omitted.

`all` is equivalent to specifying `mahalanobis` and `generalized`.

`format(%`*fmt*`)` specifies the matrix display format. The default is `format(%9.0g)`.

Options

`priors(`*priors*`)` specifies the group prior probabilities and affects only the output of the `generalized` option. By default, *priors* is determined from `e(grouppriors)`. See *Options for predict* for the *priors* specification.

Remarks and examples

stata.com

The `predict` and `estat` commands after `discrim qda` help in exploring the QDA model. See [MV] **discrim estat** for details of the `estat` subcommands common to all `discrim` subcommands. Here we illustrate some of these common `estat` subcommands along with `estat covariance`, `estat correlations`, and `estat grdistances` that are specific to `discrim qda`.

▷ Example 1: Group covariances and correlations and out-of-sample prediction

Everitt and Dunn (2001, 269) show data for male Egyptian skulls from the early and late predynastic epochs. Ten observations from each epoch are provided. Four measurements were taken of each skull: `x1`, maximum breadth; `x2`, basibregmatic height; `x3`, basialveolar length; and `x4`, nasal height. All measurements were in millimeters. Everitt and Dunn obtained the data from Manly (2005).

We perform a quadratic discriminant analysis on this dataset and demonstrate the use of `estat` and `predict`.

```
. use http://www.stata-press.com/data/r15/skulls
(Egyptian Skulls)
. discrim qda x1 x2 x3 x4, group(predynastic)
Quadratic discriminant analysis
Resubstitution classification summary
```

Key
Number Percent

True predynastic	Classified early	late	Total
early	9 90.00	1 10.00	10 100.00
late	3 30.00	7 70.00	10 100.00
Total	12 60.00	8 40.00	20 100.00
Priors	0.5000	0.5000	

What kind of covariance structure do the two groups have? If they are similar to one another, we might wish to switch to using LDA (see [MV] **discrim lda**) instead of QDA. `estat covariance` displays the covariance matrices for our two groups.

```
. estat covariance
Group covariance matrices
  predynastic: early
```

	x1	x2	x3	x4
x1	40.32222			
x2	7.188889	15.34444		
x3	13.18889	-7.322222	36.9	
x4	16.1	8.077778	-2.144444	11.43333

```
  predynastic: late
```

	x1	x2	x3	x4
x1	43.12222			
x2	-4.966667	38.98889		
x3	9.388889	6.611111	10.27778	
x4	5.211111	12.74444	4.388889	9.122222

There appear to be differences, including differences in sign between some of the elements of the covariance matrices of the two groups. How substantial are these differences? The `estat correlations` command displays the correlation matrices for the groups. The `p` option requests that p-values be presented with the correlations.

```
. estat correlations, p
Group correlation matrices
  predynastic: early
```

Key
Correlation Two-sided p-value

	x1	x2	x3	x4
x1	1.00000			
x2	0.28901 0.41800	1.00000		
x3	0.34192 0.33353	-0.30772 0.38707	1.00000	
x4	0.74984 0.01251	0.60986 0.06119	-0.10440 0.77409	1.00000

```
predynastic: late
```

Key
Correlation Two-sided p-value

	x1	x2	x3	x4
x1	1.00000			
x2	-0.12113 0.73889	1.00000		
x3	0.44598 0.19640	0.33026 0.35133	1.00000	
x4	0.26274 0.46331	0.67577 0.03196	0.45327 0.18830	1.00000

Few of the correlations in the two matrices are statistically significant. We are less sure of the apparent differences between the covariance structures for these two groups.

Let's press forward anyway. Everitt and Dunn (2001, 269) ask for the prediction for an unknown skull. We input the unknown observation and then use `predict` to obtain the classification and probabilities of group membership for the observation.

```
. input
     predyna~c    x1    x2    x3    x4
 21. . 127 129 95 51
 22. end
. predict grp
(option classification assumed; group classification)
. predict pr1 pr2, pr
. label values grp epoch
. list x* grp pr1 pr2 in 21
```

	x1	x2	x3	x4	grp	pr1	pr2
21.	127	129	95	51	late	.3654425	.6345575

This skull is classified by our QDA model as belonging to the late predynastic epoch with probability 0.63.

`estat list` could also be used to obtain this same information; see [MV] **discrim estat**.

```
. estat list in 21, varlist
```

	Data				Classification		Probabilities	
Obs.	x1	x2	x3	x4	True	Class.	early	late
21	127	129	95	51		late	0.3654	0.6346

We could use `predict` and `estat` to explore other aspects of this QDA model, including leave-one-out (LOO) classifications, probabilities, classification tables, and error-rate estimates.

◁

Example 2: Mahalanobis and generalized distances between groups

Example 1 of [MV] **discrim qda** performs a QDA on the apple tree rootstock data found in Andrews and Herzberg (1985, 357–360). We now demonstrate the use of the `estat grdistances` command for examining the squared Mahalanobis distances and the squared generalized distances between the rootstock groups.

```
. use http://www.stata-press.com/data/r15/rootstock, clear
(Table 6.2 Rootstock Data -- Rencher (2002))
. discrim qda y1 y2 y3 y4, group(rootstock) notable
. estat grdistances, all
Mahalanobis squared distances between groups
```

rootstock	rootstock 1	2	3	4	5
1	0	18.37241	7.89988	1.622808	14.78843
2	42.19008	0	5.489408	14.08784	1.502462
3	36.81811	1.908369	0	6.406024	15.48121
4	2.281963	14.77928	6.742393	0	25.72128
5	33.70858	1.855704	4.617755	16.34139	0
6	3.860684	17.32868	12.5828	11.24491	3.49512

rootstock	rootstock 6
1	9.152132
2	30.45472
3	72.60112
4	29.01146
5	20.50925
6	0

```
Generalized squared distances between groups
```

rootstock	rootstock 1	2	3	4	5
1	-17.89946	2.47128	-9.577605	-14.60611	-1.796629
2	24.29063	-15.90113	-11.98808	-2.141072	-15.0826
3	18.91866	-13.99276	-17.47749	-9.822891	-1.103849
4	-15.61749	-1.121858	-10.73509	-16.22892	9.136221
5	15.80913	-14.04543	-12.85973	.1124762	-16.58506
6	-14.03877	1.427543	-4.894681	-4.984005	-13.08994

rootstock	rootstock 6
1	-7.241371
2	14.06121
3	56.20761
4	12.61796
5	4.115752
6	-16.3935

Both tables are nonsymmetric. For QDA the Mahalanobis distance depends on the covariance of the reference group. The Mahalanobis distance for group i (the rows in the tables above) to group j (the columns in the tables above) will use the covariance matrix of group j in determining the distance. The generalized distance also factors in the prior probabilities for the groups, and so the diagonal elements are not zero and the entries can be negative. In either matrix, the smaller the number, the closer the groups.

Stored results

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

Matrices

`r(Rho_#)`	group # correlation matrix
`r(P_#)`	two-sided p-values for group # correlations

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

Matrices

`r(S_#)`	group # covariance matrix

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

Matrices

`r(sqdist)`	Mahalanobis squared distances between group means (`mahalanobis` only)
`r(gsqdist)`	generalized squared distances between group means (`generalized` only)

Methods and formulas

See *Methods and formulas* of [MV] **discrim qda** for background on what is produced by `predict` and `estat grdistances`. See [MV] **discrim estat** for more information on `estat classtable`, `estat errorrate`, `estat grsummarize`, and `estat list`.

References

Andrews, D. F., and A. M. Herzberg, ed. 1985. *Data: A Collection of Problems from Many Fields for the Student and Research Worker*. New York: Springer.

Everitt, B. S., and G. Dunn. 2001. *Applied Multivariate Data Analysis*. 2nd ed. London: Arnold.

Manly, B. F. J. 2005. *Multivariate Statistical Methods: A Primer*. 3rd ed. London: Chapman & Hall.

Rencher, A. C., and W. F. Christensen. 2012. *Methods of Multivariate Analysis*. 3rd ed. Hoboken, NJ: Wiley.

Also see

[MV] **discrim qda** — Quadratic discriminant analysis

[MV] **discrim estat** — Postestimation tools for discrim

[MV] **discrim** — Discriminant analysis

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