discrim logistic postestimation — Postestimation tools for discrim logistic

Postestimation commands predict Remarks and examples Reference Also see

Postestimation commands

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

Command	Description
`estat classtable`	classification table
`estat errorrate`	classification error-rate estimation
`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 membership, probabilities of group membership, etc.

*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 and probabilities.

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

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` and specify either one or `e(N_groups)` new variables with `pr`. `group()` is not allowed with `classification`.

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.

`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`.

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.

Remarks and examples

Classifications and probabilities after `discrim logistic` are obtained with the `predict` command. The common `estat` subcommands after `discrim` are also available for producing classification tables, error-rate tables, classification listings, and group summaries; see [MV] **discrim estat**.

▷ Example 1: Error rates and predictions for logistic discriminant analysis

Continuing with our logistic discriminant analysis of the senility dataset of Morrison (2005), introduced in example 1 of [MV] **discrim logistic**, we illustrate the use of the `estat errorrate` postestimation command.

```
. use https://www.stata-press.com/data/r19/senile
(Senility WAIS subtest scores)
. discrim logistic sim pc, group(sf) priors(proportional) notable nolog
. estat errorrate, pp
Error rate estimated from posterior probabilities
```

Error rate	sf No-SF	SF	Total
Stratified	.0305051	.5940575	.168518
Unstratified	.0305051	.5940575	.168518
Priors	.755102	.244898	

We specified the pp option to obtain the posterior probability–based error-rate estimates. The stratified and unstratified estimates are identical because proportional priors were used. The estimates were based on proportional priors because the logistic discriminant analysis model used proportional priors and we did not specify the `priors()` option in our call to `estat errorrate`.

The error-rate estimate for the senile-factor group is much higher than for the no-senile-factor group.

What error-rate estimates would we obtain with equal group priors?

```
. estat errorrate, pp priors(equal)
Error rate estimated from posterior probabilities
```

Error rate	sf No-SF	SF	Total
Stratified	.2508207	.2069481	.2288844
Unstratified	.06308	.4289397	.2460098
Priors	.5	.5	

Stratified and unstratified estimates are now different. This happens when group sizes have a different proportion from that of the prior probabilities.

Morrison (2005, 231) shows a classification of the subjects where, if the estimated probability of belonging to the senile-factor group is less than 0.35, he classifies the subject to the no-senile-factor group; if the probability is more than 0.66, he classifies the subject to the senile-factor group; and if the probability is between those extremes, he classifies the subject to an uncertain group.

We can use `predict` to implement this same strategy. The `pr` option requests probabilities. Because the model was estimated with proportional prior probabilities, the prediction, by default, will also be based on proportional prior probabilities.

```
. predict prob0 prob1, pr
. generate newgrp = 1
. replace newgrp = 0 if prob1 <= 0.35
(38 real changes made)
. replace newgrp = 2 if prob1 >= 0.66
(5 real changes made)
. label define newgrp 0 "No-SF" 1 "Uncertain" 2 "SF"
. label values newgrp newgrp
. tabulate sf newgrp
```

Senile-factor diagnosis	newgrp No-SF	Uncertain	SF	Total
No-SF	33	4	0	37
SF	5	2	5	12
Total	38	6	5	49

Six observations are placed in the uncertain group.

Reference

Morrison, D. F. 2005. *Multivariate Statistical Methods*. 4th ed. Belmont, CA: Duxbury.

Also see

[MV] **discrim logistic** — Logistic discriminant analysis

[MV] **discrim** — Discriminant analysis

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

[U] 20 Estimation and postestimation commands