**discrim logistic** — Logistic discriminant analysis

<table>
<thead>
<tr>
<th>Description</th>
<th>Quick start</th>
<th>Menu</th>
<th>Syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Options</td>
<td>Remarks and examples</td>
<td>Stored results</td>
<td>Methods and formulas</td>
</tr>
<tr>
<td>References</td>
<td>Also see</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Description**

discrim logistic performs logistic discriminant analysis. See [MV] discrim for other discrimination commands.

**Quick start**

Logistic discriminant analysis of v1, v2, v3, and v4 for groups defined by catvar

discrim logistic v1 v2 v3 v4, group(catvar)

As above, but use prior probabilities that are proportional to group size

discrim logistic v1 v2 v3 v4, group(catvar) ///
priors(proportional)

As above, but suppress iteration log

discrim logistic v1 v2 v3 v4, group(catvar) ///
priors(proportional) nolog

**Menu**

Statistics > Multivariate analysis > Discriminant analysis > Logistic
Syntax

discrim logistic varlist [if] [in] [weight], group(groupvar) [options]

<table>
<thead>
<tr>
<th>options</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>group(groupvar)</em></td>
<td>variable specifying the groups</td>
</tr>
<tr>
<td>priors(priors)</td>
<td>group prior probabilities</td>
</tr>
<tr>
<td>ties(ties)</td>
<td>how ties in classification are to be handled</td>
</tr>
<tr>
<td>notable</td>
<td>suppress resubstitution classification table</td>
</tr>
<tr>
<td>no log</td>
<td>display or suppress the mlogit log-likelihood iteration log; default is to display</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>priors</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>equal</td>
<td>equal prior probabilities; the default</td>
</tr>
<tr>
<td>proportional</td>
<td>group-size-proportional prior probabilities</td>
</tr>
<tr>
<td>matname</td>
<td>row or column vector containing the group prior probabilities</td>
</tr>
<tr>
<td>matrix_exp</td>
<td>matrix expression providing a row or column vector of the group prior probabilities</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ties</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>missing</td>
<td>ties in group classification produce missing values; the default</td>
</tr>
<tr>
<td>random</td>
<td>ties in group classification are broken randomly</td>
</tr>
<tr>
<td>first</td>
<td>ties in group classification are set to the first tied group</td>
</tr>
</tbody>
</table>

* *group() is required.

statsby and xi are allowed; see [U] 11.1.10 Prefix commands.

fweights are allowed; see [U] 11.1.6 weight.

See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.

Options

**Model**

*group(groupvar)* is required and specifies the name of the grouping variable. *groupvar* must be a numeric variable.

priors(priors) specifies the prior probabilities for group membership. The following priors are allowed:

- priors(equal) specifies equal prior probabilities. This is the default.
- 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. The following ties are allowed:

- ties(missing) specifies that ties in group classification produce missing values. This is the default.
- 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.

Reporting

notable suppresses the computation and display of the resubstitution classification table.

log and nolog specify whether to display the mlogit log-likelihood iteration log. The iteration log is displayed by default unless you used set iterlog off to suppress it; see set iterlog in [R] set iter.

Remarks and examples

Albert and Lesaffre (1986) explain that logistic discriminant analysis is a partially parametric method falling between parametric discrimination methods such as LDA and QDA (see [MV] discrim lda and [MV] discrim qda) and nonparametric discrimination methods such as $k$th-nearest-neighbor (KNN) discrimination (see [MV] discrim knn). Albert and Harris (1987) provide a good explanation of logistic discriminant analysis. Instead of making assumptions about the distribution of the data within each group, logistic discriminant analysis is based on the assumption that the likelihood ratios of the groups have an exponential form; see Methods and formulas. Multinomial logistic regression provides the basis for logistic discriminant analysis; see [R] mlogit. Multinomial logistic regression can handle binary and continuous regressors, and hence logistic discriminant analysis is also appropriate for binary and continuous discriminating variables.

Example 1: A two-group logistic discriminant analysis

Morrison (2005, 443–445) provides data on 12 subjects with a senile-factor diagnosis and 37 subjects with a no-senile-factor diagnosis. The data consist of the Wechsler Adult Intelligence Scale (WAIS) subtest scores for information, similarities, arithmetic, and picture completion. Morrison (2005, 231) performs a logistic discriminant analysis on the two groups, using the similarities and picture completion scores as the discriminating variables.
. use https://www.stata-press.com/data/r16/senile
(Senility WAIS subtest scores)
. discrim logistic sim pc, group(sf) priors(proportional)
Iteration 0:  log likelihood = -27.276352
Iteration 1:  log likelihood = -19.531198
Iteration 2:  log likelihood = -19.036702
Iteration 3:  log likelihood = -19.018973
Iteration 4:  log likelihood = -19.018928
Logistic discriminant analysis
Resubstitution classification summary

<table>
<thead>
<tr>
<th>Key</th>
<th>Number</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>True sf</td>
<td>No-SF</td>
<td>SF</td>
</tr>
<tr>
<td>No-SF</td>
<td>37</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>100.00</td>
<td>0.00</td>
</tr>
<tr>
<td>SF</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>50.00</td>
<td>50.00</td>
</tr>
<tr>
<td>Total</td>
<td>43</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>87.76</td>
<td>12.24</td>
</tr>
<tr>
<td>Priors</td>
<td>0.7551</td>
<td>0.2449</td>
</tr>
</tbody>
</table>

We specified the priors(proportional) option to obtain proportional prior probabilities for our logistic classification. These results match those of Morrison (2005, 231), though he does not state that his results are based on proportional prior probabilities. If you change to equal prior probabilities you obtain different classification results.

Which observations were misclassified? estat list with the misclassified option shows the six misclassified observations and the estimated probabilities.

. estat list, misclassified varlist

<table>
<thead>
<tr>
<th>Obs.</th>
<th>Data</th>
<th>Classification</th>
<th>Probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sim</td>
<td>pc</td>
<td>True</td>
</tr>
<tr>
<td>38</td>
<td>5</td>
<td>8</td>
<td>SF</td>
</tr>
<tr>
<td>41</td>
<td>7</td>
<td>9</td>
<td>SF</td>
</tr>
<tr>
<td>44</td>
<td>9</td>
<td>8</td>
<td>SF</td>
</tr>
<tr>
<td>46</td>
<td>7</td>
<td>6</td>
<td>SF</td>
</tr>
<tr>
<td>48</td>
<td>10</td>
<td>3</td>
<td>SF</td>
</tr>
<tr>
<td>49</td>
<td>12</td>
<td>10</td>
<td>SF</td>
</tr>
</tbody>
</table>

* indicates misclassified observations

See example 1 of [MV] discrim logistic postestimation for more postestimation analysis with this logistic discriminant analysis.
Example 2 of [MV] discrim knn introduces a head measurement dataset with six discriminating variables and three groups; see Rencher and Christensen (2012, 290–292). We now apply discrim logistic to see how well the logistic model can discriminate between the groups.

```
. use https://www.stata-press.com/data/r16/head
    (Table 8.3 Head measurements, Rencher and Christensen (2012))
. discrim logistic wdim circum fbeye eyehd earhd jaw, group(group)
Iteration 0:  log likelihood =  -98.875106
Iteration 1:  log likelihood =  -60.790737
Iteration 2:  log likelihood =  -53.746934
Iteration 3:  log likelihood =  -51.114631
Iteration 4:  log likelihood =  -50.249426
Iteration 5:  log likelihood =  -50.081199
Iteration 6:  log likelihood =  -50.072248
Iteration 7:  log likelihood =  -50.072216
```

Logistic discriminant analysis
Resubstitution classification summary

<table>
<thead>
<tr>
<th>True group</th>
<th>Classified</th>
<th></th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>high school</td>
<td>college</td>
<td>nonplayer</td>
<td></td>
</tr>
<tr>
<td>high school</td>
<td>27</td>
<td>2</td>
<td>1</td>
<td></td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>90.00</td>
<td>6.67</td>
<td>3.33</td>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td>college</td>
<td>1</td>
<td>20</td>
<td>9</td>
<td></td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>3.33</td>
<td>66.67</td>
<td>30.00</td>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td>nonplayer</td>
<td>2</td>
<td>8</td>
<td>20</td>
<td></td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>6.67</td>
<td>26.67</td>
<td>66.67</td>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td></td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>33.33</td>
<td>33.33</td>
<td>33.33</td>
<td></td>
<td>100.00</td>
</tr>
</tbody>
</table>

Priors: 0.3333

The counts on the diagonal of the resubstitution classification table are similar to those obtained by discrim knn (see example 2 of [MV] discrim knn) and discrim lda (see example 1 of [MV] candisc), whereas discrim qda seems to have classified the nonplayer group more accurately (see example 3 of [MV] discrim estat).
 Stored results

discrim logistic stores the following in e():

Scalars
- e(N) number of observations
- e(N_groups) number of groups
- e(k) number of discriminating variables
- e(ibaseout) base outcome number

Macros
- e(cmd) discrim
- e(subcmd) logistic
- e(cmdline) command as typed
- e(groupvar) name of group variable
- e(grouplabels) labels for the groups
- e(varlist) discriminating variables
- e(dropped) variables dropped because of collinearity
- e(wtype) weight type
- e(wexp) weight expression
- e(title) title in estimation output
- e(ties) how ties are to be handled
- e(properties) b noV
- e(estat_cmd) program used to implement estat
- e(predict) program used to implement predict
- e(marginsnotok) predictions disallowed by margins

Matrices
- e(b) coefficient vector
- e(groupcounts) number of observations for each group
- e(grouppriors) prior probabilities for each group
- e(groupvalues) numeric value for each group

Functions
- e(sample) marks estimation sample

Methods and formulas

Let \( g \) be the number of groups, \( n_i \) the number of observations for group \( i \), and \( q_i \) the prior probability for group \( i \). Let \( \mathbf{x} \) denote an observation measured on \( p \) discriminating variables. For consistency with the discriminant analysis literature, \( \mathbf{x} \) will be a column vector, though it corresponds to a row in your dataset. Let \( f_i(\mathbf{x}) \) represent the density function for group \( i \), and let \( P(\mathbf{x}|G_i) \) denote the probability of observing \( \mathbf{x} \) conditional on belonging to group \( i \). Denote the posterior probability of group \( i \) given observation \( \mathbf{x} \) as \( P(G_i|\mathbf{x}) \). With Bayes’s theorem, we have

\[
P(G_i|\mathbf{x}) = \frac{q_i f_i(\mathbf{x})}{\sum_{j=1}^{g} q_j f_j(\mathbf{x})}
\]

Substituting \( P(\mathbf{x}|G_i) \) for \( f_i(\mathbf{x}) \), we have

\[
P(G_i|\mathbf{x}) = \frac{q_i P(\mathbf{x}|G_i)}{\sum_{j=1}^{g} q_j P(\mathbf{x}|G_j)}
\]
Dividing both the numerator and denominator by \( P(\mathbf{x} | G_g) \), we can express this as

\[
P(\mathbf{G}_i | \mathbf{x}) = \frac{q_i L_{ig}(\mathbf{x})}{\sum_{j=1}^{g} q_j L_{jg}(\mathbf{x})}
\]

where \( L_{ig}(\mathbf{x}) = P(\mathbf{x} | \mathbf{G}_i) / P(\mathbf{x} | G_g) \) is the likelihood ratio of \( \mathbf{x} \) for groups \( i \) and \( g \).

This formulation of the posterior probability allows easy insertion of the Multinomial logistic model into the discriminant analysis framework. The multinomial logistic model expresses \( L_{ig}(\mathbf{x}) \) in a simple exponential form

\[
L_{ig}(\mathbf{x}) = \exp(a_{0i} + a_i' \mathbf{x})
\]

see Albert and Harris (1987, 117). Logistic discriminant analysis uses \texttt{mlogit} to compute the likelihood ratios, \( L_{ig}(\mathbf{x}) \), and hence the posterior probabilities \( P(\mathbf{G}_i | \mathbf{x}) \); see \texttt{[R] mlogit}. However, \texttt{mlogit} and \texttt{predict} after \texttt{mlogit} assume proportional prior probabilities. \texttt{discrim logistic} assumes equal prior probabilities unless you specify the \texttt{priors(proportional)} option.

**References**


**Also see**

\texttt{[MV] discrim logistic postestimation} — Postestimation tools for discrim logistic

\texttt{[MV] discrim} — Discriminant analysis

\texttt{[R] logistic} — Logistic regression, reporting odds ratios

\texttt{[R] mlogit} — Multinomial (polytomous) logistic regression

\texttt{[U] 20 Estimation and postestimation commands}