

**discrim lda** — Linear discriminant analysis

<a href="#">Description</a>	<a href="#">Quick start</a>	<a href="#">Menu</a>	<a href="#">Syntax</a>
<a href="#">Options</a>	<a href="#">Remarks and examples</a>	<a href="#">Stored results</a>	<a href="#">Methods and formulas</a>
<a href="#">References</a>	<a href="#">Also see</a>		

## Description

`discrim lda` performs linear discriminant analysis. See [MV] [discrim](#) for other discrimination commands.

If you want canonical linear discriminant results displayed, see [MV] [candisc](#).

## Quick start

Linear discriminant analysis of `v1`, `v2`, `v3`, and `v4` for groups defined by `catvar`  
`discrim lda v1 v2 v3 v4, group(catvar)`

As above, but use prior probabilities proportional to group size  
`discrim lda v1 v2 v3 v4, group(catvar) priors(proportional)`

Display the leave-one-out and the resubstitution classification tables  
`discrim lda v1 v2 v3 v4, group(catvar) lootable`

As above, but suppress the resubstitution classification table  
`discrim lda v1 v2 v3 v4, group(catvar) lootable notable`

## Menu

Statistics > Multivariate analysis > Discriminant analysis > Linear (LDA)

## Syntax

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

<i>options</i>	Description
<b>Model</b>	
* <u>group</u> ( <i>groupvar</i> )	variable specifying the groups
<u>priors</u> ( <i>priors</i> )	group prior probabilities
<u>ties</u> ( <i>ties</i> )	how ties in classification are to be handled
<b>Reporting</b>	
<u>notable</u>	suppress resubstitution classification table
<u>lootable</u>	display leave-one-out classification table
<b>priors</b>	
<b>Description</b>	
<u>equal</u>	equal prior probabilities; the default
<u>proportional</u>	group-size-proportional prior probabilities
<i>matname</i>	row or column vector containing the group prior probabilities
<i>matrix_exp</i>	matrix expression providing a row or column vector of the group prior probabilities
<b>ties</b>	
<b>Description</b>	
<u>missing</u>	ties in group classification produce missing values; the default
<u>random</u>	ties in group classification are broken randomly
<u>first</u>	ties in group classification are set to the first tied group

\*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.

`lootable` displays the leave-one-out classification table.

## Remarks and examples

[stata.com](http://stata.com)

`discrim lda` computes the same things as `candisc`, but `candisc` displays more information. The same information can be displayed after `discrim lda` by using the `estat` suite of commands; see [MV] [discrim lda postestimation](#).

Remarks are presented under the following headings:

*Introduction*

*Descriptive LDA*

*Predictive LDA*

*A classic example*

## Introduction

Linear discriminant analysis (LDA) was developed by different researchers, [Fisher \(1936\)](#) and [Mahalanobis \(1936\)](#), starting with different approaches to the problem of discriminating between groups. [Kshirsagar and Arseven \(1975\)](#), [Green \(1979\)](#), and [Williams \(1982\)](#) demonstrate the mathematical relationship between Fisher's linear discriminant functions and the classification functions from the Mahalanobis approach to LDA; see [Rencher \(1998, 239\)](#).

Fisher's approach to LDA forms the basis of descriptive LDA but can be used for predictive LDA. The Mahalanobis approach to LDA more naturally handles predictive LDA, allowing for prior probabilities and producing estimates of the posterior probabilities. The Mahalanobis approach to LDA also extends to quadratic discriminant analysis (QDA); see [MV] [discrim qda](#).

## Descriptive LDA

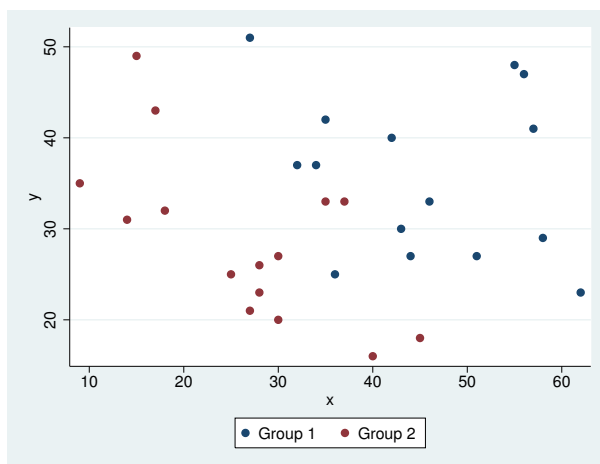
[Fisher \(1936\)](#) approached linear discriminant analysis by seeking the linear combination of the discriminating variables that provides maximal separation between the groups (originally two groups, but later extended to multiple groups). Maximal separation of groups is determined from an eigen analysis of  $\mathbf{W}^{-1}\mathbf{B}$ , where  $\mathbf{B}$  is the between-group sum-of-squares and cross-products (SSCP) matrix, and  $\mathbf{W}$  is the within-group SSCP matrix. The eigenvalues and eigenvectors of  $\mathbf{W}^{-1}\mathbf{B}$  provide what are called Fisher's linear discriminant functions.

The first linear discriminant function is the eigenvector associated with the largest eigenvalue. This first discriminant function provides a linear transformation of the original discriminating variables into one dimension that has maximal separation between group means. The eigenvector associated with the second-largest eigenvalue is the second linear discriminant function and provides a dimension uncorrelated with (but usually not orthogonal to) the first discriminant function. The second discriminant function provides the maximal separation of groups in a second dimension. The third discriminant function provides the maximum separation of groups in a third dimension.

## ▷ Example 1

Two groups measured on two variables illustrate Fisher's approach to linear discriminant analysis.

```
. use http://www.stata-press.com/data/r14/twogroup
(Two Groups)
```



Splitting the two groups on the basis of  $y$  or  $x$  alone would leave a great deal of overlap between the groups. By eye it appears that a line with a slope of about  $-1$  could be positioned between the two groups with only a few of the observations falling on the wrong side of the line.

Fisher's approach to LDA helps us find the best separating line.

```
. discrim lda y x, group(group) notable
```

`discrim lda` computes the information needed for both a predictive and descriptive linear discriminant analysis. We requested `notable`, which suppressed the display of the resubstitution classification table. We will examine this feature of `discrim lda` when we discuss predictive LDA. The descriptive features of LDA are available through postestimation commands.

```
. estat loadings, unstandardized
```

Canonical discriminant function coefficients

	function1
y	.0862145
x	.0994392
_cons	-6.35128

Fisher's linear discriminant functions provide the basis for what are called the canonical discriminant functions; see *Methods and formulas*. The canonical discriminant function coefficients are also called unstandardized loadings because they apply to the unstandardized discriminating variables ( $x$  and  $y$ ). Because we have only two groups, there is only one discriminant function. From the coefficients or loadings of this discriminant function, we obtain a one-dimensional projection of the data that gives maximal separation between the two groups relative to the spread within the groups. The `estat loadings` postestimation command displayed these loadings; see [\[MV\] discrim lda postestimation](#). After `estat loadings`, the unstandardized loadings are available in matrix `r(L_unstd)`. We take these values and determine the equation of the separating line between the groups and a line perpendicular to the separating line.

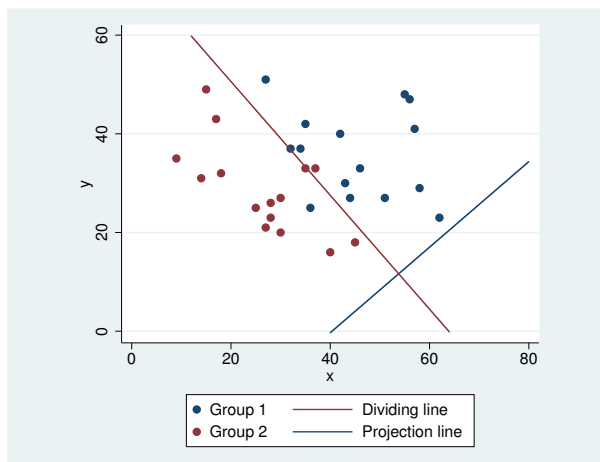
The unstandardized canonical discriminant function coefficients indicate that

$$0 = 0.0862145y + 0.0994392x - 6.35128$$

which in standard  $y = mx + b$  form is

$$y = -1.1534x + 73.6684$$

which is the dividing line for classifying observations into the two groups for this LDA. A line perpendicular to this dividing line has slope  $-1 / -1.153 = 0.867$ . The following graph shows the data with this dividing line and a perpendicular projection line.



Another way of thinking about the discriminant function is that it projects the data from the original two dimensions down to one dimension—the line perpendicular to the dividing line. Classifications are based on which side of the separating line the observations fall.

Researchers often wish to know which of the discriminating variables is most important or helpful in discriminating between the groups. They want to examine the standardized loadings—the loadings that apply to standardized variables.

```
. estat loadings, standardized
Standardized canonical discriminant function coefficients
```

	function1
y	.7798206
x	1.057076

These coefficients or loadings apply to  $x$  and  $y$  that have been standardized using the pooled within-group covariance matrix. The loading for  $x$  is larger than that for  $y$ , indicating that it contributes more to the discrimination between the groups. Look back at the scatterplot to see that there is more separation between the groups in the  $x$  dimension than the  $y$  dimension. See [\[MV\] discrim lda postestimation](#) for more details of the `estat loadings` command.

Some researchers prefer to examine what are called structure coefficients.

```
. estat structure
Canonical structure
```

	function1
y	.3146309
x	.7138982

The `estat structure` command provides structure coefficients, which measure the correlation between each discriminating variable and the discriminant function; see [MV] **discrim lda postestimation**. Here the canonical structure coefficient for `x` is larger than that for `y`, leading to the same conclusion as with standardized loadings. There is disagreement in the literature concerning the use of canonical structure coefficients versus standardized loadings; see [Rencher and Christensen \(2012, 300–301\)](#) and [Huberty \(1994, 262–264\)](#).

◀

In addition to loading and structure coefficients, there are other descriptive LDA features available after `discrim lda`. These include canonical correlations and tests of the canonical correlations, classification functions, scree plots, loading plots, score plots, and various group summaries; see [MV] **discrim lda postestimation**.

If your main interest is in descriptive LDA, you may find the `candisc` command of interest; see [MV] **candisc**. `discrim lda` and `candisc` differ only in their default output. `discrim lda` shows classification tables. `candisc` shows canonical correlations, standardized coefficients (loadings), structure coefficients, and more. All the features found in [MV] **discrim lda postestimation** are available for both commands.

## Predictive LDA

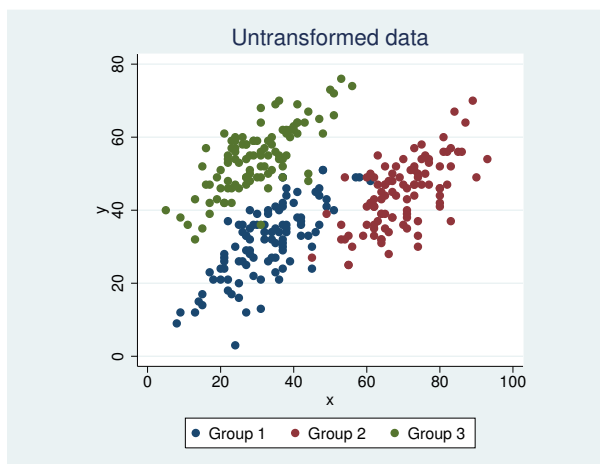
Another approach to linear discriminant analysis starts with the assumption that the observations from each group are multivariate normal with the groups having equal covariance matrices but different means. [Mahalanobis \(1936\)](#) distance plays an important role in this approach. An observation with unknown group membership is classified as belonging to the group with smallest Mahalanobis distance between the observation and group mean. Classification functions for classifying observations of unknown group membership can also be derived from this approach to LDA and formulas for the posterior probability of group membership are available.

As shown in [Methods and formulas](#), Mahalanobis distance can be viewed as a transformation followed by Euclidean distance. Group membership is assigned based on the Euclidean distance in this transformed space.

### ▷ Example 2

We illustrate the Mahalanobis transformation and show some of the features of predictive discriminant analysis with a simple three-group example dataset.

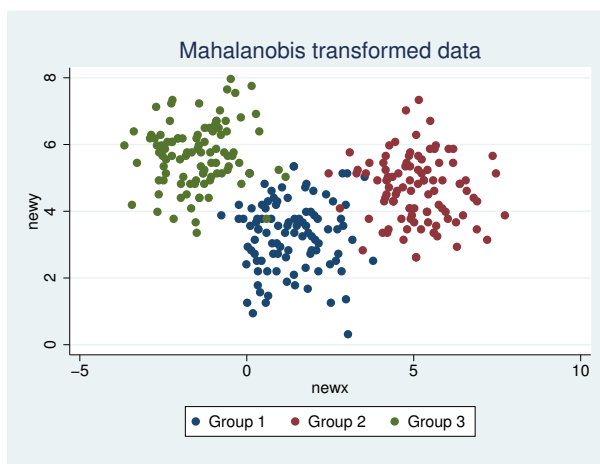
```
. use http://www.stata-press.com/data/r14/threegroup  
(Three Groups)
```



These three groups appear to have similar covariance structure—showing a positive correlation between  $x$  and  $y$ . There is some overlap of the three groups, but general identification of the groups does not appear difficult by human eye.

If we were to apply Euclidean distance for classifying this untransformed data, we would misclassify some observations that clearly should not be misclassified when judged by eye. For example, in the graph above, the observations from group 3 that have  $y$  values below 40 (found in the lower left of the group 3 cloud of points) are closer in Euclidean distance to the center of group 1.

The following graph shows the Mahalanobis-transformed data.



With the transformed data, using Euclidean distance between observations and group means works well.

Now let's see how well `discrim lda` can distinguish between the groups. By default, a resubstitution classification table is presented. The rows correspond to the known group and the columns to the grouping as assigned by the `discrim` model. The word `resubstitution` is used because the same observations that built the model are being classified by the model.

```
. discrim lda y x, group(group)
Linear discriminant analysis
Resubstitution classification summary
```

Key	
Number	Percent

True group	Classified			Total
	1	2	3	
1	93	4	3	100
	93.00	4.00	3.00	100.00
2	3	97	0	100
	3.00	97.00	0.00	100.00
3	3	0	97	100
	3.00	0.00	97.00	100.00
Total	99	101	100	300
	33.00	33.67	33.33	100.00
Priors	0.3333	0.3333	0.3333	

For these 300 observations, group 1 had 93 observations correctly classified, 4 observations misclassified into group 2, and 3 observations misclassified into group 3. Group 2 had 3 observations misclassified into group 1 and 97 observations correctly classified. Group 3 had 3 observations misclassified into group 1 and 97 observations correctly classified.

Generally, resubstitution classification tables give an overly optimistic view of how well you would classify an unknown observation. Leave-one-out (LOO) classification tables provide a more realistic assessment for classification success. With this 300-observation dataset, the LOO classification table gives the same results. We could see the LOO classification table by requesting it at estimation, by requesting it at replay, or by using the `estat classtable` command.



We now list the misclassified observations.

```
. estat list, varlist misclassified
```

Obs.	Data		Classification		Probabilities		
	y	x	True	Class.	1	2	3
19	49	37	1	3 *	0.2559	0.0000	0.7441
29	49	57	1	2 *	0.4245	0.5750	0.0005
47	49	37	1	3 *	0.2559	0.0000	0.7441
55	24	45	1	2 *	0.4428	0.5572	0.0000
70	48	61	1	2 *	0.0661	0.9339	0.0000
74	49	58	1	2 *	0.3041	0.6957	0.0003
92	37	22	1	3 *	0.3969	0.0000	0.6031
143	27	45	2	1 *	0.6262	0.3738	0.0000
161	39	49	2	1 *	0.8026	0.1973	0.0001
185	49	54	2	1 *	0.7782	0.2187	0.0030
238	48	44	3	1 *	0.8982	0.0017	0.1001
268	50	44	3	1 *	0.7523	0.0009	0.2469
278	36	31	3	1 *	0.9739	0.0000	0.0261

\* indicates misclassified observations

The posterior probabilities for each displayed observation for each of the three groups is presented along with the true group and the classified group. The observation number is also shown. We added the discriminating variables *x* and *y* to the list with the `varlist` option. By default, `estat list` would list all the observations. The `misclassified` option restricts the list to those observations that were misclassified.

With `predict` we could generate classification variables, posterior probabilities, Mahalanobis squared distances from observations to group means, classification function scores (see [Methods and formulas](#)), and more. Fifteen `estat` commands provide more predictive and descriptive tools after `discrim lda`; see [\[MV\] discrim lda postestimation](#).

◀

## A classic example

We use the iris data from Fisher's (1936) pioneering LDA article to demonstrate the `discrim lda` command.

### ▶ Example 3

Fisher obtained the iris data from [Anderson \(1935\)](#). The data consist of four features measured on 50 samples from each of three iris species. The four features are the length and width of the sepal and petal. The three species are *Iris setosa*, *Iris versicolor*, and *Iris virginica*. [Morrison \(2005, app. B.2\)](#) is a modern source of the data.

```
. use http://www.stata-press.com/data/r14/iris
(Iris data)
```

Running `discrim lda` produces the resubstitution classification table.

```
. discrim lda seplen sepwid petlen petwid, group(iris)
```

Linear discriminant analysis

Resubstitution classification summary

Key		Classified			Total
Number	Percent	setosa	versicolor	virginica	
True iris					
setosa	50 100.00	0 0.00	0 0.00	0 0.00	50 100.00
versicolor	0 0.00	48 96.00	2 4.00	2 4.00	50 100.00
virginica	0 0.00	1 2.00	49 98.00	1 2.00	50 100.00
Total	50 33.33	49 32.67	51 34.00	51 34.00	150 100.00
Priors	0.3333	0.3333	0.3333	0.3333	

One *Iris virginica* observation was misclassified as a *versicolor*, two *Iris versicolor* observations were misclassified as *virginica*, and no *Iris setosa* observations were misclassified in our resubstitution classification.

Which observations were misclassified?

```
. estat list, misclassified
```

Obs.	Classification		Probabilities		
	True	Class.	setosa	versicolor	virginica
71	versicol	virginic *	0.0000	0.2532	0.7468
84	versicol	virginic *	0.0000	0.1434	0.8566
134	virginic	versicol *	0.0000	0.7294	0.2706

\* indicates misclassified observations

Postestimation command `estat list` shows that observations 71, 84, and 134 were misclassified and shows the estimated posterior probabilities for the three species for the misclassified observations.

We now examine the canonical discriminant functions for this LDA. The number of discriminant functions will be one fewer than the number of groups or will be the number of discriminating variables, whichever is less. With four discriminating variables and three species, we will have two discriminant functions. `estat loadings` displays the discriminant coefficients or loadings.

```
. estat loadings, unstandardized standardized
```

```
Canonical discriminant function coefficients
```

	function1	function2
seplen	-.8293776	-.0241021
sepwid	-1.534473	-2.164521
petlen	2.201212	.9319212
petwid	2.81046	-2.839188
_cons	-2.105106	6.661473

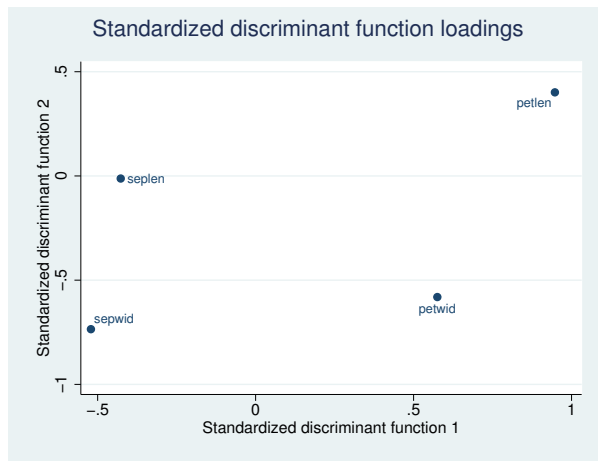
```
Standardized canonical discriminant function coefficients
```

	function1	function2
seplen	-.4269548	-.0124075
sepwid	-.5212417	-.7352613
petlen	.9472572	.4010378
petwid	.5751608	-.5810399

We requested the display of both unstandardized and standardized loadings. The two unstandardized discriminant functions provide linear combinations of the `seplen`, `sepwid`, `petlen`, and `petwid` discriminating variables—producing two new dimensions. The standardized canonical discriminant function coefficients indicate the relative importance and relationship between the discriminating variables and the discriminant functions. The first discriminant function compares `seplen` and `sepwid`, which have negative standardized coefficients, to `petlen` and `petwid`, which have positive standardized coefficients. The second discriminant function appears to be contrasting the two width variables from the two length variables, though this is not as distinct of a difference as found in the first discriminant function because the `seplen` variable in the second standardized discriminant function is close to zero.

Understanding the composition of the discriminant functions is aided by plotting the coefficients. `loadingplot` graphs the discriminant function coefficients (loadings); see [MV] [discrim lda postestimation](#) and [MV] [scoreplot](#).

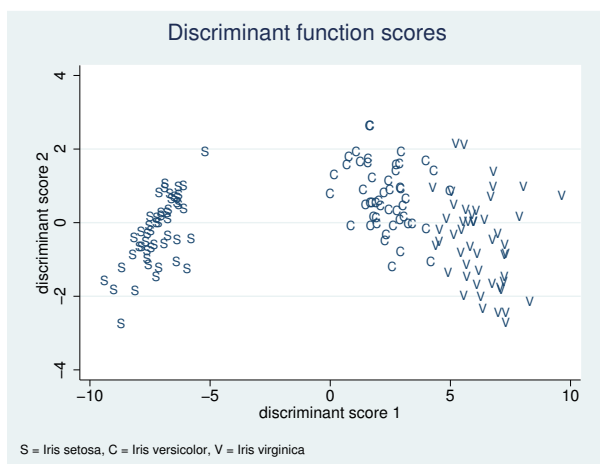
```
. loadingplot
```



We now show a scatterplot of our three species projected onto the two dimensions of our discriminant solution. The `scoreplot` command takes care of most of the graphing details for us; see [MV] [discrim lda postestimation](#) and [MV] [scoreplot](#). However, by default, `scoreplot` uses the full value labels for the three iris species and the resulting graph is busy. The `iris` dataset has two label languages

predefined. The default label language has the full value labels. The other predefined label language is called `oneletter`, and it uses a one-letter code as value labels for the three iris species. The `label language` command will switch between these two label languages; see [D] [label language](#). We also use the `msymbol(i)` graph option so that the points will not be displayed—only the one-letter value labels will be displayed for each observation.

```
. label language oneletter
. scoreplot, msymbol(i)
>       note("S = Iris setosa, C = Iris versicolor, V = Iris virginica")
```



The *Iris setosa* are well separated from the other two species. *Iris versicolor* and *Iris virginica* show some overlap.

See [example 1](#) of [MV] [discrim estat](#) and [examples 6, 7, and 8](#), of [MV] [discrim lda postestimation](#) for more examples of what can be produced after `discrim lda` for this iris dataset.

## Stored results

discrim lda stores the following in `e()`:

### Scalars

<code>e(N)</code>	number of observations
<code>e(N_groups)</code>	number of groups
<code>e(k)</code>	number of discriminating variables
<code>e(f)</code>	number of nonzero eigenvalues

### Macros

<code>e(cmd)</code>	discrim
<code>e(subcmd)</code>	lda
<code>e(cmdline)</code>	command as typed
<code>e(groupvar)</code>	name of group variable
<code>e(grouplabels)</code>	labels for the groups
<code>e(varlist)</code>	discriminating variables
<code>e(wtype)</code>	weight type
<code>e(wexp)</code>	weight expression
<code>e(title)</code>	title in estimation output
<code>e(ties)</code>	how ties are to be handled
<code>e(properties)</code>	no <code>b</code> no <code>V</code> eigen
<code>e(estat_cmd)</code>	program used to implement <code>estat</code>
<code>e(predict)</code>	program used to implement <code>predict</code>
<code>e(marginsnotok)</code>	predictions disallowed by <code>margins</code>

### Matrices

<code>e(groupcounts)</code>	number of observations for each group
<code>e(grouppriors)</code>	prior probabilities for each group
<code>e(groupvalues)</code>	numeric value for each group
<code>e(means)</code>	group means on discriminating variables
<code>e(SSCP_W)</code>	pooled within-group SSCP matrix
<code>e(SSCP_B)</code>	between-groups SSCP matrix
<code>e(SSCP_T)</code>	total SSCP matrix
<code>e(SSCP_W#)</code>	within-group SSCP matrix for group #
<code>e(W_eigvals)</code>	eigenvalues of <code>e(SSCP_W)</code>
<code>e(W_eigvecs)</code>	eigenvectors of <code>e(SSCP_W)</code>
<code>e(S)</code>	pooled within-group covariance matrix
<code>e(Sinv)</code>	inverse of <code>e(S)</code>
<code>e(sqrtSinv)</code>	Cholesky (square root) of <code>e(Sinv)</code>
<code>e(Ev)</code>	eigenvalues of $\mathbf{W}^{-1}\mathbf{B}$
<code>e(L_raw)</code>	eigenvectors of $\mathbf{W}^{-1}\mathbf{B}$
<code>e(L_unstd)</code>	unstandardized canonical discriminant function coefficients
<code>e(L_std)</code>	within-group standardized canonical discriminant function coefficients
<code>e(L_totalstd)</code>	total-sample standardized canonical discriminant function coefficients
<code>e(C)</code>	classification coefficients
<code>e(cmeans)</code>	unstandardized canonical discriminant functions evaluated at group means
<code>e(canstruct)</code>	canonical structure matrix
<code>e(candisc_stat)</code>	canonical discriminant analysis statistics

### Functions

<code>e(sample)</code>	marks estimation sample
------------------------	-------------------------

## Methods and formulas

Methods and formulas are presented under the following headings:

*Predictive LDA*  
*Descriptive LDA*

## Predictive LDA

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

$$P(G_i|\mathbf{x}) = \frac{q_i P(\mathbf{x}|G_i)}{\sum_{j=1}^g q_j P(\mathbf{x}|G_j)}$$

LDA assumes that the groups are multivariate normal with equal covariance matrices. Let  $\mathbf{S}$  denote the pooled within-group sample covariance matrix and  $\bar{\mathbf{x}}_i$  denote the sample mean of group  $i$ . The  $\bar{\mathbf{x}}_i$  are returned as the columns of the `e(means)` matrix. The squared Mahalanobis distance between observation  $\mathbf{x}$  and  $\bar{\mathbf{x}}_i$  is

$$D_i^2 = (\mathbf{x} - \bar{\mathbf{x}}_i)' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_i)$$

Plugging these sample estimates into the multivariate normal density gives

$$P(\mathbf{x}|G_i) = (2\pi)^{-p/2} |\mathbf{S}|^{-1/2} e^{-D_i^2/2}$$

Substituting this into the formula for  $P(G_i|\mathbf{x})$  and simplifying gives

$$P(G_i|\mathbf{x}) = \frac{q_i e^{-D_i^2/2}}{\sum_{j=1}^g q_j e^{-D_j^2/2}}$$

as the LDA posterior probability of observation  $\mathbf{x}$  belonging to group  $i$ .

Computation of Mahalanobis distance can be broken down into two steps. Step one: transform the data by using the Mahalanobis transformation. Step two: compute the Euclidean distance of the transformed data.

Let  $\mathbf{L}$  be the Cholesky factorization of  $\mathbf{S}^{-1}$  such that  $\mathbf{S}^{-1} = \mathbf{L}'\mathbf{L}$  and  $\mathbf{L}$  is lower triangular.  $\mathbf{L}$  is returned in matrix `e(sqrtSinv)`. Squared Mahalanobis distance can be expressed in terms of  $\mathbf{L}$ .

$$\begin{aligned} D_i^2 &= (\mathbf{x} - \bar{\mathbf{x}}_i)' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_i) \\ &= (\mathbf{x} - \bar{\mathbf{x}}_i)' \mathbf{L}'\mathbf{L} (\mathbf{x} - \bar{\mathbf{x}}_i) \\ &= (\mathbf{L}\mathbf{x} - \mathbf{L}\bar{\mathbf{x}}_i)' (\mathbf{L}\mathbf{x} - \mathbf{L}\bar{\mathbf{x}}_i) \\ &= (\mathbf{z} - \bar{\mathbf{z}}_i)' (\mathbf{z} - \bar{\mathbf{z}}_i) \end{aligned}$$

which is the squared Euclidean distance between  $\mathbf{z}$  and  $\bar{\mathbf{z}}_i$ . We call  $\mathbf{z} = \mathbf{L}\mathbf{x}$  the Mahalanobis transformation.

The squared Mahalanobis distance between group means is produced by `estat grdistances`; see [\[MV\] discrim lda postestimation](#).

Classification functions can be derived from the Mahalanobis formulation for LDA; see [Rencher and Christensen \(2012, 315–316\)](#) and [Huberty \(1994, 59\)](#). Let  $L_i(\mathbf{x})$  denote the linear classification function for the  $i$ th group applied to observation  $\mathbf{x}$ .  $L_i(\mathbf{x}) = \mathbf{c}_i' \mathbf{x} + c_{i0}$ , where  $\mathbf{c}_i = \bar{\mathbf{x}}_i' \mathbf{S}^{-1}$  and  $c_{i0} = -(1/2) \bar{\mathbf{x}}_i' \mathbf{S}^{-1} \bar{\mathbf{x}}_i + \ln(q_i)$ . The  $g$  linear classification functions are returned as the columns of matrix `e(C)` and through the `estat classfunction` command; see [\[MV\] discrim lda postestimation](#). An observation can be classified based on largest posterior probability or based on largest classification function score.

## Descriptive LDA

As with predictive LDA, let  $g$  be the number groups,  $n_i$  the number of training (sample) observations for group  $i$ ,  $p$  the number of discriminating variables, and  $N = \sum_{i=1}^g n_i$  the total number of observations. Also, let  $\mathbf{W}$  be the within-group sums-of-squares and cross-products (SSCP) matrix and let  $\mathbf{B}$  be the between-groups SSCP matrix. Fisher's (1936) linear discriminant functions are based on the eigenvalues and eigenvectors of  $\mathbf{W}^{-1}\mathbf{B}$ .

There are  $s = \min(g - 1, p)$  nonzero eigenvalues of  $\mathbf{W}^{-1}\mathbf{B}$ . Let  $\lambda_1, \lambda_2, \dots, \lambda_s$  denote the  $s$  eigenvalues in decreasing order. The eigenvalues are returned in `e(Ev)`. Let  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s$  denote the corresponding eigenvectors. Rencher and Christensen (2012, 289) outlines the approach for obtaining the eigenvalues and eigenvectors of the nonsymmetric  $\mathbf{W}^{-1}\mathbf{B}$  matrix. Because  $\mathbf{W}^{-1}\mathbf{B}$  is nonsymmetric, the resulting eigenvectors are not orthogonal but merely uncorrelated; see Rencher and Christensen (2012, 289). A matrix with the  $\mathbf{v}_i$  as columns is returned in `e(L_raw)`. The phrase raw coefficients is used by Klecka (1980, 22) to describe the  $\mathbf{v}$  vectors.

Any constant multiple of the eigenvector  $\mathbf{v}_i$  is still an eigenvector of  $\mathbf{W}^{-1}\mathbf{B}$  associated with eigenvalue  $\lambda_i$ . Typically,  $\mathbf{v}\mathbf{u}_i = \mathbf{v}_i\sqrt{N - g}$  are used as the eigenvectors (discriminant functions) and are called unstandardized canonical discriminant functions because they correspond to the unstandardized discriminating variables. An extra element is added to the bottom of the  $\mathbf{v}\mathbf{u}$  vectors for the constant, so that if the  $\mathbf{v}\mathbf{u}$  vectors are applied as linear combinations of the discriminating variables, the resulting variables have mean zero; see Klecka (1980, 21–23). A matrix with the  $\mathbf{v}\mathbf{u}_i$  as columns is returned in `e(L_unstd)`.

The means of the discriminating variables for each group are returned as the columns of the matrix `e(means)`. These group means multiplied by the unstandardized discriminant-function coefficients,  $\mathbf{v}\mathbf{u}_i$ , produce what are called group means on canonical variables and are returned in the matrix `e(cmeans)` and displayed with the command `estat grmeans, canonical`.

Standardized discriminant functions are obtained as  $\mathbf{v}\mathbf{s}_i = \mathbf{v}_i\sqrt{W_{ii}}$ . The  $i$ th raw eigenvector is standardized by the square root of the  $i$ th diagonal element of the  $\mathbf{W}$  matrix. These within-group standardized discriminant functions are used in assessing the importance and relationship of the original discriminating variables to the discriminant functions. A matrix with the  $\mathbf{v}\mathbf{s}_i$  as columns is returned in `e(L_std)`.

Let  $\mathbf{T}$  denote the total sample SSCP matrix. Total-sample standardized discriminant functions are computed as  $\mathbf{v}\mathbf{t}_i = \mathbf{v}_i\sqrt{T_{ii}(N - g)/(N - 1)}$ . A matrix with the  $\mathbf{v}\mathbf{t}_i$  as columns is returned in `e(L_totalstd)`. There is debate as to which of  $\mathbf{v}\mathbf{s}$  and  $\mathbf{v}\mathbf{t}$  should be used for interpretation; see Mueller and Cozad (1988), Nordlund and Nagel (1991), and Mueller and Cozad (1993).

The `estat loadings` command displays `e(L_unstd)`, `e(L_std)`, and `e(L_totalstd)`; see [MV] [discrim lda postestimation](#).

The canonical structure matrix measures the correlation between the discriminating variables and the discriminant function and is returned in matrix `e(canstruct)`. The canonical structure matrix is equal to  $\mathbf{W}\mathbf{V}$  with the  $i$ th row divided by  $\sqrt{W_{ii}}$ , where  $\mathbf{V}$  contains the  $\mathbf{v}_i$  eigenvectors as columns. Rencher and Christensen (2012, 300) warns against the use of structure coefficients for interpretation, but Huberty (1994, 262–264) contends otherwise.

The returned matrix `e(candisc_stat)` contains columns for the information shown by `estat canontest`, including the eigenvalues, canonical correlations, proportion of variance, cumulative proportion of variance, likelihood-ratio test statistics, and the corresponding  $F$  tests, degrees of freedom, and  $p$ -values. See [MV] [canon](#).

As noted in the [Introduction](#) section of *Remarks and examples*, Kshirsagar and Arseven (1975), Green (1979), and Williams (1982) demonstrate the mathematical relationship between Fisher's

linear discriminant functions (the basis for descriptive LDA) and the classification functions from the Mahalanobis approach to LDA (the basis for predictive LDA); see [Rencher \(1998, 239\)](#).

## References

- Anderson, E. 1935. The irises of the Gaspé Peninsula. *Bulletin of the American Iris Society* 59: 2–5.
- Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7: 179–188.
- Green, B. F. 1979. The two kinds of linear discriminant functions and their relationship. *Journal of Educational Statistics* 4: 247–263.
- Hilbe, J. M. 1992. [smv3: Regression-based dichotomous discriminant analysis](#). *Stata Technical Bulletin* 5: 13–17. Reprinted in *Stata Technical Bulletin Reprints*, vol. 1, pp. 131–137. College Station, TX: Stata Press.
- . 1996. [smv3.1: Discriminant analysis: An enhanced command](#). *Stata Technical Bulletin* 34: 34–36. Reprinted in *Stata Technical Bulletin Reprints*, vol. 6, pp. 190–192. College Station, TX: Stata Press.
- . 1997. [smv3.2: Enhancements to discriminant analysis](#). *Stata Technical Bulletin* 38: 26–27. Reprinted in *Stata Technical Bulletin Reprints*, vol. 7, p. 229. College Station, TX: Stata Press.
- Huberty, C. J. 1994. *Applied Discriminant Analysis*. New York: Wiley.
- Johnson, R. A., and D. W. Wichern. 2007. *Applied Multivariate Statistical Analysis*. 6th ed. Englewood Cliffs, NJ: Prentice Hall.
- Klecka, W. R. 1980. *Discriminant Analysis*. Newbury Park, CA: Sage.
- Kshirsagar, A. M., and E. Arseven. 1975. A note on the equivalency of two discrimination procedures. *American Statistician* 29: 38–39.
- Lachenbruch, P. A. 1975. *Discriminant Analysis*. New York: Hafner Press.
- Mahalanobis, P. C. 1936. On the generalized distance in statistics. *National Institute of Science of India* 12: 49–55.
- Mardia, K. V., J. T. Kent, and J. M. Bibby. 1979. *Multivariate Analysis*. London: Academic Press.
- McLachlan, G. J. 2004. *Discriminant Analysis and Statistical Pattern Recognition*. New York: Wiley.
- Morrison, D. F. 2005. *Multivariate Statistical Methods*. 4th ed. Belmont, CA: Duxbury.
- Mueller, R. O., and J. B. Cozad. 1988. Standardized discriminant coefficients: Which variance estimate is appropriate? *Journal of Educational Statistics* 13: 313–318.
- . 1993. Standardized discriminant coefficients: A rejoinder. *Journal of Educational Statistics* 18: 108–114.
- Nordlund, D. J., and R. W. Nagel. 1991. Standardized discriminant coefficients revisited. *Journal of Educational Statistics* 16: 101–108.
- Rencher, A. C. 1998. *Multivariate Statistical Inference and Applications*. New York: Wiley.
- Rencher, A. C., and W. F. Christensen. 2012. *Methods of Multivariate Analysis*. 3rd ed. Hoboken, NJ: Wiley.
- Tabachnick, B. G., and L. S. Fidell. 2007. *Using Multivariate Statistics*. 5th ed. Boston: Allyn and Bacon.
- Williams, B. K. 1982. A simple demonstration of the relationship between classification and canonical variates analysis. *American Statistician* 36: 363–365.

## Also see

- [MV] [discrim lda postestimation](#) — Postestimation tools for `discrim lda`
- [MV] [discrim](#) — Discriminant analysis
- [MV] [candisc](#) — Canonical linear discriminant analysis
- [U] [20 Estimation and postestimation commands](#)