

User Guide: Discriminate
P-value estimation for Discriminant Function Analyses

Discriminate comes in 2 parts. The Matlab file `discriminate.m` performs the calculations, and the Excel file `DataImport.xls` is the interface for entering data. Advanced knowledge of Matlab is not required for using this software, although your copy of Matlab must include the Statistical Toolbox.

Data need to be formatted in the usual way for DFA: a matrix of predictor variable values, with a column for each variable and a row for each sample. Coupled with the data matrix is a column vector with a Group ID label for each sample (i.e. each sample is assigned to Group 1, Group 2, etc., so the ID labels will be 1, 2, etc.). Each row in the data matrix will be paired with a corresponding entry in the Group ID vector.

To use the software:

1. Input data into `DataImport.xls`.

Paste the raw data underneath the column headings for “Predictor Variables.” Use as many columns as necessary. DFA requires multivariate normality, so data should be transformed accordingly prior to using this software.

To the left of the data, paste the group labels under the column heading “Group IDs.”

There must be a group ID for each data entry. Samples must have a value for each response variable (no missing values).

2. Select number of randomizations.

Enter the desired number of randomizations into the field indicated. We recommend at least 1000 randomizations.

3. Place `DataImport.xls` and `discriminate.m` into the same directory or folder.

4. Open Matlab and select the directory from step (3) as the Current Directory.

5. At the prompt in the Command Window, type:

```
discriminate();
```

6. The program will return 7 values:

the null expectation for the jackknife reclassification success
the standard deviation for the null expectation
the 95%, 50%, and 5% quantiles for the null distribution of reclassification success values
the actual jackknife reclassification success for your dataset
the *P*-value (one-tailed) for your jackknife reclassification success

The P -value is for the null hypothesis that the observed reclassification success is no better than that expected by random chance. Note that it is an estimate and will become more precise with a larger number of randomizations. The minimum value returned for P is $1/(\# \text{ of randomizations})$.

Modifications to the Matlab code:

By default, the Matlab code uses a linear DFA with uniform prior probabilities. If your analysis calls for a different type of DFA, you may modify `discriminate.m` to reflect this. The program uses Matlab's built-in DFA tool, 'classify.' To change the DFA procedure, alter the input arguments for the 'classify' function on line 82 of the program code. See Matlab's Help documentation for `classify.m` for more information.

Dec. 14, 2005

J. Wilson White

Department of Ecology, Evolution, and Marine Biology

University of California

Santa Barbara, CA 93106

w_white@lifesci.ucsb.edu