# Linear Discriminant Analysis (LDA)

Multivariate Analysis

Mathematical Background, Worked Example, Tool Usage

English

**Wolfgang Boehme**

wolfgang.boehme@economic-modeling.de

www.economic-modeling.de

8 November 2026

The documentation on this website was developed with the support of the AI Claude (Anthropic).

# Table of Contents

**1 Introduction and Application Areas** **2**

**2 Mathematical Background** **2**
2.1 Scatter Matrices . . . 2
2.2 Discriminant Functions . . . 2
2.3 Discriminant Scores . . . 3
2.4 Standardised Coefficients . . . 3
2.5 Canonical Correlation . . . 3
2.6 Wilks' Lambda . . . 3
2.7 Bartlett Test of Significance . . . 4
2.8 Leave-One-Out Cross-Validation (LOO-CV) . . . 4

**3 Worked Example: Balance-Sheet Ratios** **4**
3.1 Data . . . 4
3.2 Step 1: Group Statistics . . . 4
3.3 Step 2: Discriminant Functions . . . 5
3.4 Step 3: Standardised Discriminant Coefficients . . . 5
3.5 Step 4: Group Centroids in Discriminant Space . . . 6
3.6 Step 5: Classification and Cross-Validation . . . 6

**4 Using the Interactive LDA Tool** **6**
4.1 Data Input . . . 6
4.2 Parameter Selection . . . 7
4.3 Results Tabs . . . 7
4.4 Typical Workflow . . . 8

**5 Formula Reference** **9**

**6 Further Reading** **9**

## Introduction and Application Areas

**Linear Discriminant Analysis** (LDA) is a *supervised* classification method: unlike cluster analysis, group memberships of the observations are known. LDA finds **discriminant functions** – linear combinations of the variables – that separate the groups as well as possible.

Dual objectives of LDA:

- **Descriptive:** Which variables contribute to group separation?
- **Classification:** How is a new observation assigned to a group?

**Typical application areas:**

- Credit scoring (good / bad debtors)
- Medical diagnosis (ill / healthy)
- Market segmentation with known target groups
- Quality control (conforming / non-conforming parts)
- Species identification in biology and archaeology

## Mathematical Background

### Scatter Matrices

Let $\mathbf{X} \in \mathbb{R}^{n \times p}$ with $g$ known groups, $n_i$ observations per group and group mean $\bar{\mathbf{x}}_i$.

**Within-group scatter matrix $\mathbf{S}_W$:**

$$\mathbf{S}_W = \sum_{i=1}^{g} \sum_{j \in C_i} (\mathbf{x}_j - \bar{\mathbf{x}}_i)(\mathbf{x}_j - \bar{\mathbf{x}}_i)^\top \tag{1}$$

**Between-group scatter matrix $\mathbf{S}_B$:**

$$\mathbf{S}_B = \sum_{i=1}^{g} n_i (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^\top \tag{2}$$

### Discriminant Functions

LDA maximises the Rayleigh quotient criterion:

$$\max_{\mathbf{w}} \frac{\mathbf{w}^\top \mathbf{S}_B \mathbf{w}}{\mathbf{w}^\top \mathbf{S}_W \mathbf{w}} \tag{3}$$

This leads to the generalised eigenvalue problem:

$$\mathbf{S}_B \, \mathbf{w} = \lambda \, \mathbf{S}_W \, \mathbf{w} \quad \Longleftrightarrow \quad \mathbf{S}_W^{-1} \mathbf{S}_B \, \mathbf{w} = \lambda \, \mathbf{w} \tag{4}$$

The solutions $\mathbf{w}_1, \mathbf{w}_2, \ldots$ are the **discriminant coefficients** of discriminant functions DF1, DF2, ...

Since $\mathbf{S}_W^{-1}\mathbf{S}_B$ is not symmetric, the tool solves the problem via the Cholesky decomposition $\mathbf{S}_W = \mathbf{L}\mathbf{L}^\top$: the symmetric matrix $\mathbf{C} = \mathbf{L}^{-1}\mathbf{S}_B\mathbf{L}^{-\top}$ has the same eigenvalues $\lambda_k$; its eigenvectors $\mathbf{u}_k$ (Jacobi method) yield $\mathbf{w}_k = \sqrt{n-g}\,\mathbf{L}^{-\top}\mathbf{u}_k$. With this normalisation the discriminant scores have a pooled within-group variance of 1.

The maximum number of meaningful discriminant functions is $\min(p,\ g-1)$.

## Discriminant Scores

$$d_{ik} = \mathbf{w}_k^\top \mathbf{x}_i = \sum_{j=1}^{p} w_{kj}\, x_{ij} \tag{5}$$

**Classification rule:** Observation $\mathbf{x}_i$ is assigned to the group $g^*$ whose centroid $\bar{d}_g$ lies closest in discriminant space (owing to the normalisation, the Euclidean distance equals the Mahalanobis distance), adjusted by the prior probabilities $\pi_g$:

$$g^* = \arg\min_{g} \Big[ \sum_{k=1}^{K} (d_{ik} - \bar{d}_{gk})^2 - 2\ln \pi_g \Big] \tag{6}$$

With equal priors ($\pi_g = 1/g$) the correction term vanishes; otherwise $\pi_g = n_g/n$.

## Standardised Coefficients

$$w_{kj}^{\text{std}} = w_{kj} \cdot s_j^{(W)}, \qquad s_j^{(W)} = \sqrt{\frac{(\mathbf{S}_W)_{jj}}{n-g}} \tag{7}$$

Scaling by the pooled within-group standard deviation $s_j^{(W)}$ makes the coefficients comparable regardless of measurement scale.

## Canonical Correlation

$$r_k^* = \sqrt{\frac{\lambda_k}{1+\lambda_k}} \tag{8}$$

$r_k^*$ is the correlation between discriminant scores DF$k$ and the group indicator variables – a measure of discriminatory power.

## Wilks' Lambda

Wilks' Lambda $\Lambda$ is the central goodness-of-fit measure of LDA:

$$\Lambda = \prod_{k=1}^{K} \frac{1}{1+\lambda_k} = \frac{|\mathbf{S}_W|}{|\mathbf{S}_W + \mathbf{S}_B|} \tag{9}$$

$\Lambda = 0$: perfect separation; $\Lambda = 1$: no separation.

### Bartlett Test of Significance

$$\chi^2 \approx -\left(n - 1 - \frac{p+g}{2}\right) \ln \Lambda, \qquad \mathrm{df} = p(g-1) \tag{10}$$

### Leave-One-Out Cross-Validation (LOO-CV)

For an unbiased estimate of classification accuracy:

1. Remove observation $i$ from the training dataset.
2. Estimate LDA on the remaining $n-1$ observations.
3. Classify observation $i$.
4. Repeat for all $n$ observations.

## Worked Example: Balance-Sheet Ratios

### Data

The same 24 synthetic companies as in Handbook 1 are used (full data table there) with six ratios: $X_1$ equity ratio, $X_2$ quick ratio, $X_3$ asset coverage II, $X_4$ return on sales, $X_5$ return on assets, $X_6$ sales growth (all in %). Unlike Handbooks 1–3, the **known group membership** is used here: $g = 3$ rating classes A, B, C with $n_g = 8$ companies each (U01–U08: A, U09–U16: B, U17–U24: C). Computed with z-standardised variables and equal priors (tool defaults); the means in Step 1 are given in original units, as in the tool.

**Goal of the discriminant analysis:** determine which ratios best separate the rating classes and find a rule for assigning new companies to a rating class.

### Step 1: Group Statistics

Table 1: Group means (in %)

| Group | $X_1$ Equity | $X_2$ Liqu. | $X_3$ AC II | $X_4$ ROS | $X_5$ ROA | $X_6$ Growth |
|---|---|---|---|---|---|---|
| Rating A ($n = 8$) | 45.14 | 127.63 | 125.13 | 5.26 | 7.26 | 5.20 |
| Rating B ($n = 8$) | 34.13 | 99.38 | 96.75 | 8.55 | 11.16 | 11.71 |
| Rating C ($n = 8$) | 28.00 | 84.38 | 88.88 | 1.88 | 4.60 | −0.36 |
| Total ($n = 24$) | 35.75 | 103.79 | 103.58 | 5.23 | 7.68 | 5.52 |

A companies are characterised by high stability, B companies by high profitability; C companies have the lowest values on all ratios.

### Step 2: Discriminant Functions

There are $\min(p,\, g-1) = \min(6,\, 2) = 2$ discriminant functions.

Table 2: Eigenvalues and explained variance

| Function | Eigenvalue $\lambda$ | $r_k^*$ | Variance | Cumulative |
|---|---|---|---|---|
| DF1 | 8.003 | 0.943 | 71.8 % | 71.8 % |
| DF2 | 3.146 | 0.871 | 28.2 % | 100.0 % |

Example: $r_1^* = \sqrt{8.003/9.003} = 0.943$; variance share of DF1 $= 8.003/(8.003 + 3.146) = 71.8\,\%$.

Wilks' Lambda:

$$\Lambda = \frac{1}{1+8.003} \cdot \frac{1}{1+3.146} = 0.1111 \cdot 0.2412 = 0.0268 \quad \text{(very good separation)},$$

Bartlett test: $\chi^2 = -\left(24 - 1 - \frac{6+3}{2}\right) \ln 0.0268 = 18.5 \cdot 3.620 = 66.96$, df $= 6 \cdot 2 = 12$, $p < 0.001$.

### Step 3: Standardised Discriminant Coefficients

Table 3: Standardised coefficients

| Variable | DF1 | DF2 |
|---|---|---|
| $X_1$ Equity ratio | −0.038 | 0.193 |
| $X_2$ Quick ratio | 0.160 | 0.342 |
| $X_3$ Asset coverage | 0.069 | **0.726** |
| $X_4$ Return on sales | **0.748** | −0.184 |
| $X_5$ Return on assets | **0.488** | −0.281 |
| $X_6$ Sales growth | 0.339 | 0.149 |

**Interpretation:**

- **DF1** (71.8 % of separation) is a **profitability function**: return on sales, return on assets and growth carry the largest weights.
- **DF2** (28.2 %) is a **stability function**: asset coverage and quick ratio dominate.
- The equity ratio has only small weights in both functions, although it varies clearly between the groups. Reason: it is highly correlated with the quick ratio and asset coverage; standardised coefficients measure the contribution *in addition* to the other variables.

Discriminant analysis thus finds the same two dimensions as PCA and factor analysis – but weighted so that the rating classes are separated as well as possible. The signs of discriminant functions are arbitrary; mirroring a function does not change the results.

### Step 4: Group Centroids in Discriminant Space

Table 4: Group centroids

| Group | DF1 centroid | DF2 centroid |
|---|---|---|
| Rating A | 0.22 | 2.34 |
| Rating B | 3.13 | −1.29 |
| Rating C | −3.35 | −1.05 |

DF1 separates the highly profitable B companies from the C companies (distance 6.5); DF2 sets the A companies with their high stability apart from both other groups. In the group plot (DF1 vs. DF2) the three classes form clearly separated clouds.

### Step 5: Classification and Cross-Validation

Table 5: Confusion matrix (training data)

| | Predicted | | |
|---|---|---|---|
| Actual | A | B | C |
| Rating A | **8** | 0 | 0 |
| Rating B | 0 | **8** | 0 |
| Rating C | 0 | 0 | **8** |

**Accuracy (resubstitution): 100 %** (24 out of 24 correct).

This rate is too optimistic, because each company is classified with a function whose estimation it contributed to. **Leave-one-out cross-validation** (activate via the checkbox in the tool) yields **91.7 %** (22 out of 24): two A companies are misclassified when they do not contribute to the estimation:

- **U03** → rating C: below-average stability and returns (cf. cluster analysis, Handbook 3).
- **U07** → rating B: very high profitability ratios (return on sales 8.8 %, return on assets 9.5 %, growth 10.1 %) – a profitability profile like that of a B company.

The difference of 8.3 percentage points between resubstitution and LOO-CV is typical for $n = 24$ and $p = 6$ and shows why cross-validation is recommended for small samples.

## Using the Interactive LDA Tool

### Data Input

A distinctive feature of the LDA tool: the **GRP column** on the far left of the manual input table contains group labels (text or numbers, e.g. “Setosa”, “Good”, “1”). Every row must contain a group label to be included in the analysis.

Pre-loaded: Balance-sheet ratios (3 groups rating A/B/C, $n = 24$, $p = 6$). Further examples: Iris (3 groups, $n = 15$, $p = 4$), Credit scoring (2 groups, $n = 21$, $p = 4$) and Market segments (3 groups, $n = 24$, $p = 4$). For each example, an explanatory box above the input table describes the data, the goal of the analysis and the expected result; it disappears when "Clear" is clicked or a CSV file is loaded.

## Parameter Selection

**Standardisation**
Z-transformation recommended. If all variables share the same unit, "None" may be selected – the unstandardised coefficients are then directly interpretable.

**Number of discriminant functions**
*Automatic:* $\min(p, g-1)$ functions. Manual: 1 to 3 functions (at most $\min(p, g-1)$); classification and charts then use only these functions. Wilks' $\Lambda$ is always computed over all functions.

**Group priors**
Checkbox ticked: equal priors (all groups equally probable). Standard and recommended when no prior probabilities are known. Checkbox cleared: priors proportional to group sizes ($n_g/n$).

**LOO cross-validation**
Checkbox: Computes the leave-one-out accuracy (computationally more intensive). Recommended for small samples ($n < 50$).

## Results Tabs

**Overview**
Key figures ($n$, groups, discriminant functions, accuracy) + Wilks' Lambda table with significance assessment + DF eigenvalue table with canonical correlations + table of group means (in original units).

**Coefficients**
Standardised and unstandardised discriminant coefficients per function. Group centroids in discriminant space.

**Classification**
Confusion matrix with colour coding (diagonal green, errors red) + overall accuracy + individual classification table for each observation.

**Charts**
*Group plot (DF1 vs. DF2):* Scatter plot of all observations and group centroids (diamonds) in discriminant space. *Variance bar:* Explained variance per discriminant function. *Coefficient plot:* Bar chart of standardised absolute coefficients for DF1.

**Interpretation**
Automatic assessment text: Wilks' Lambda, Bartlett test, most important discriminators, LOO-CV comparison.

**Export**
CSV: coefficients, classification, discriminant scores. PNG: group plot, coefficient bar chart.

### Typical Workflow

1. Load data with GRP column (each group $\geq 2$ observations).
2. Leave standardisation set to z-transformation.
3. Activate the LOO-CV checkbox (for $n < 100$).
4. Click **"Calculate LDA"**.
5. Overview tab: check Wilks' $\Lambda$ and Bartlett $p$-value.
6. Coefficients tab: which variables discriminate most strongly?
7. Classification tab: analyse the confusion matrix – which groups are confused?
8. Charts tab: group plot – are the groups separable in DF1–DF2 space?
9. LOO-CV vs. training: a large difference indicates overfitting.

**Important notes:**

- Each group must contain at least 2 observations.
- Rule of thumb: $n \geq 5 \cdot p$ per group for stable LDA.
- Assumptions: multivariate normality and homogeneity of covariance matrices (homoscedasticity) per group.
- If homoscedasticity is violated: use Quadratic Discriminant Analysis (QDA).
- The accuracy reported by the tool is a *resubstitution* estimate – it overestimates the true accuracy; LOO-CV provides a less biased estimate.
- Singular $\mathbf{S}_W$ when $n < p$ or perfect multicollinearity: use more observations or fewer variables.

## Formula Reference

Table 6: Discriminant analysis formulae

| Name | Formula |
|---|---|
| Within scatter matrix | $\mathbf{S}_W = \sum_i \sum_{j \in C_i} (\mathbf{x}_j - \bar{\mathbf{x}}_i)(\mathbf{x}_j - \bar{\mathbf{x}}_i)^\top$ |
| Between scatter matrix | $\mathbf{S}_B = \sum_i n_i (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^\top$ |
| Eigenvalue problem | $\mathbf{S}_B \mathbf{w} = \lambda \mathbf{S}_W \mathbf{w}$ |
| Std. coefficient | $w_{kj} \sqrt{(\mathbf{S}_W)_{jj}/(n-g)}$ |
| Discriminant score | $d_{ik} = \mathbf{w}_k^\top \mathbf{x}_i$ |
| Canonical correlation | $r_k^* = \sqrt{\lambda_k/(1+\lambda_k)}$ |
| Wilks' Lambda | $\Lambda = \prod_k 1/(1+\lambda_k)$ |
| Bartlett $\chi^2$ | $-(n-1-(p+g)/2)\,\ln\Lambda$, $\mathrm{df} = p(g-1)$ |
| Classification | $g^* = \arg\min_g \left[ \sum_k (d_{ik} - \bar{d}_{gk})^2 - 2\ln \pi_g \right]$ |

## Further Reading


- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2), 179–188.
- Backhaus, K. et al. (2021). *Multivariate Analysemethoden*, 16th ed. Springer Gabler.
- McLachlan, G. J. (2004). *Discriminant Analysis and Statistical Pattern Recognition*. Wiley.