Diskriminant Analysis (DA)

Hypothesis Testing

(a) Interpretation of differences - Canonical Discriminant Analysis

(aa) To what extent it is possible to distinguish defined groups of objects based on the characters available.

(ab) Which of these characters contribute the most to this distinction.

(b) Identification of objects - Classification Discriminant Analysis

Derivation of one or more equations for the purpose of identifying objects.

Data Requirements:

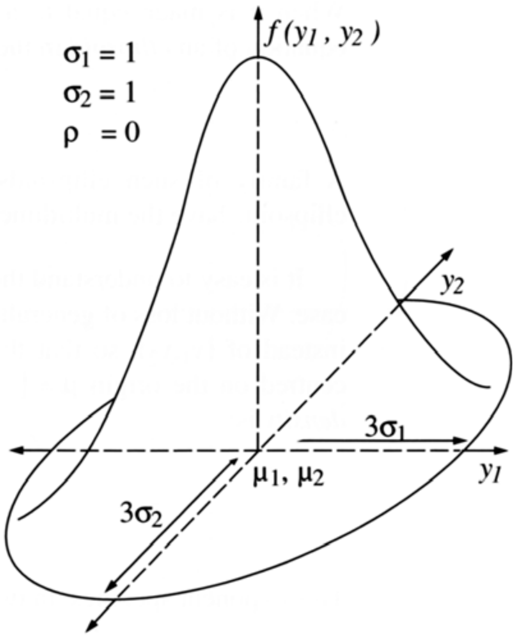
(a) Quantitative or binary characters

(b) None of the characters should be a linear combination of another character or characters

(c) It is not possible to simultaneously use two or more highly correlated characters

(d) The covariance matrices for individual groups must be approximately similar

(e) The characters describing each group should meet the requirement of multivariate normal distribution



For the number of groups (*g*), the number of characters (*p*), the number of objects in groups, and the total number of objects in the analysis (*n*) in discriminant analyses, the following must hold:

(a) There must be at least two groups of objects: $g \geq 2$;

(b) Each group must contain at least 2 objects;

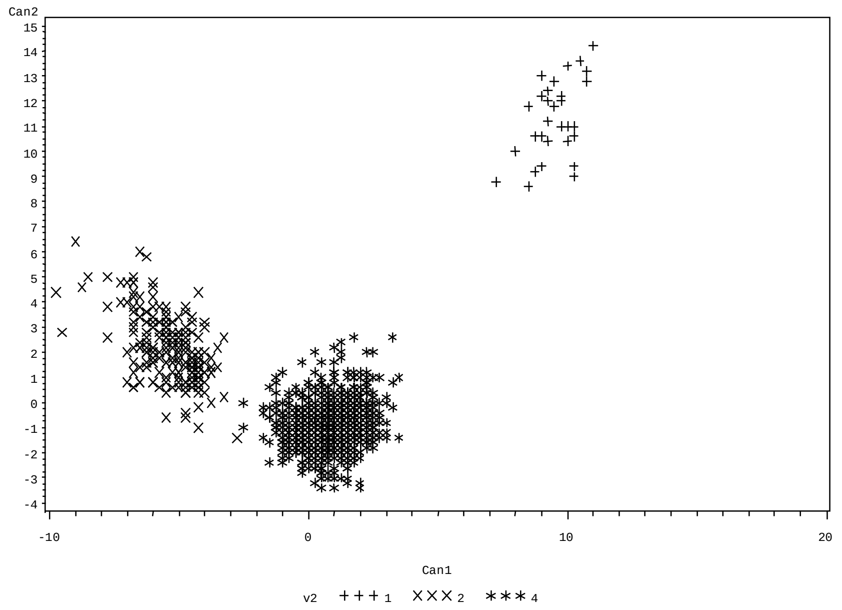
(c) The number of characters used in the analysis must be less than the number of objects reduced by the number of groups: $0 < p < (n-g)$;

(d) No character should be constant in any group.

Canonical Discriminant Analysis, Canonical Variates Analysis – CDA

Allows observing relationships between objects in a space defined by canonical axes.

An ordination procedure that maximizes differences between groups.



Canonical Discriminant Analysis, Canonical Variates Analysis – CDA

Canonical diskriminant function

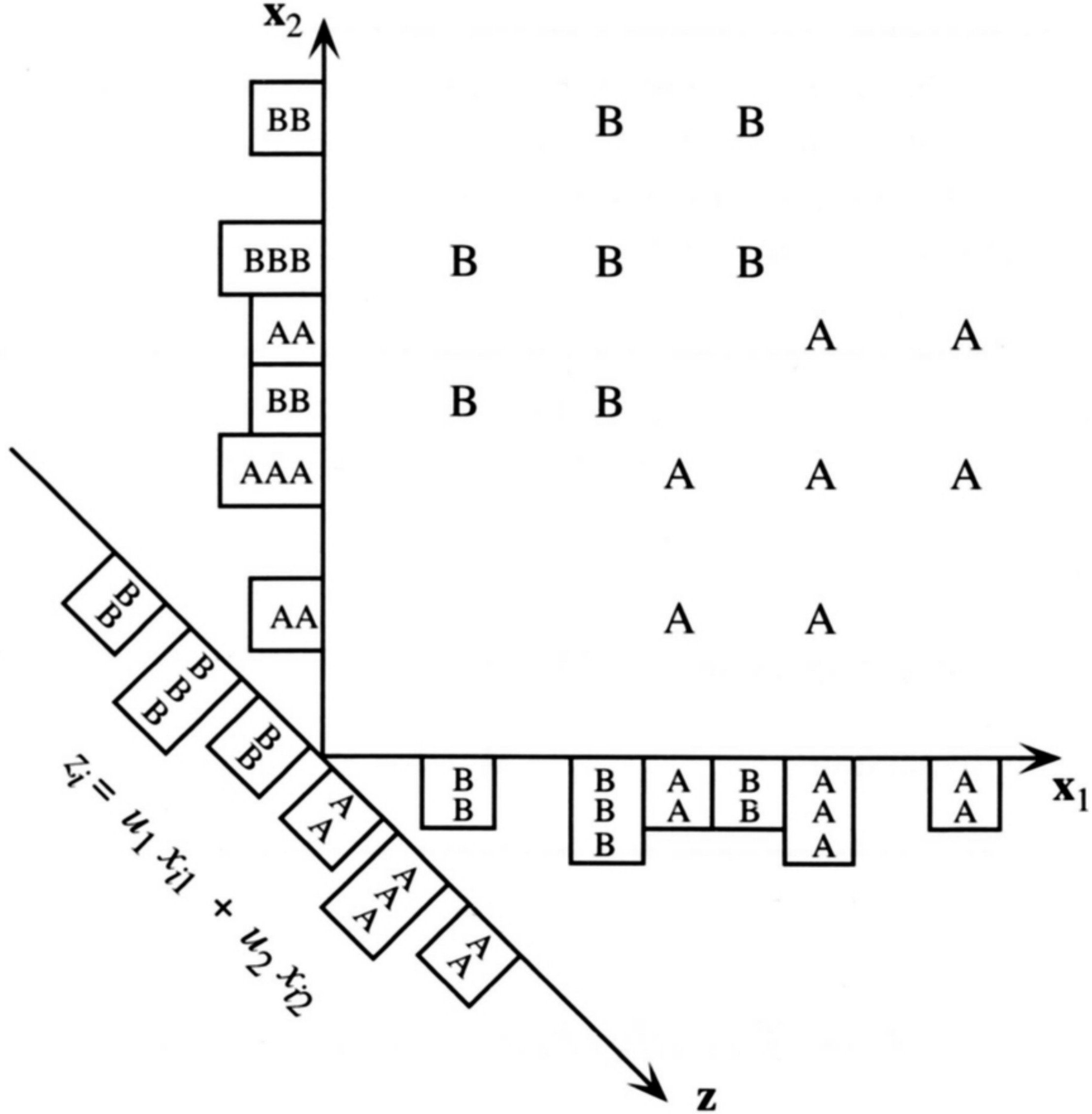
$$f_{km} = a_0 + a_1 x_{1km} + a_2 x_{2km} + \ldots + a_p x_{pkm},$$

f_{km} = the value (score) of the canonical discriminant function for case *m* in group *k*;

x_{ikm} = the value of the discriminant characteristic x_i for case *m* in group *k*.

a_i = coefficient of diskriminant function ($i = 0, 1 \ldots, p$);

The coefficients (*a*) for the first function are derived so that the group centroids (centers of gravity, means) are maximally distant (in terms of Mahalanobis distance). The coefficients calculated for the second function must further maximize the differences between group centroids, and simultaneously, the values of both functions must not be correlated.



x_2
x_1
z
$z_i = u_1 x_{i1} + u_2 x_{i2}$

	PCA, PCoA, NMDS	DA
Predefined groups	no	yes
Explanation of maximum variation	total	between groups
Character weighting	no	yes

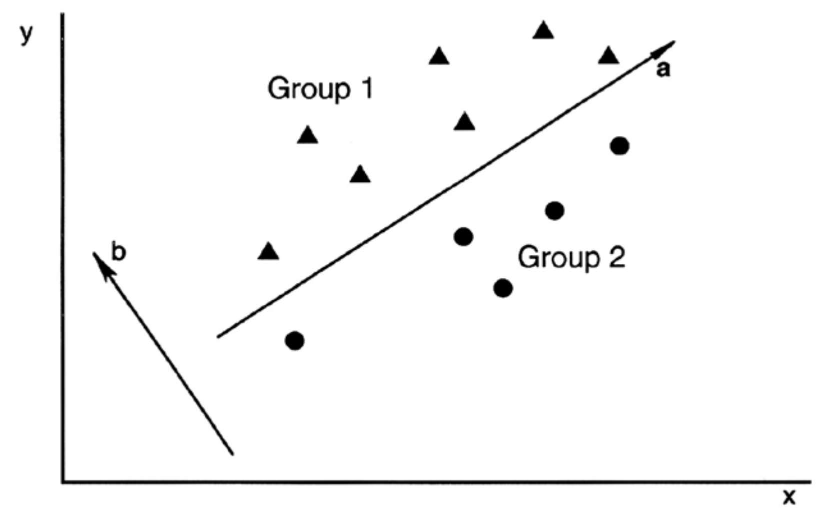


Figure 7.22. Comparison of the underlying ideas in PCA and CVA by an artificial example with two original dimensions. Component 1 (**a**) coincides with the main trend of variation in the entire sample, whereas canonical variate 1 (**b**, there is only one in this case) explains the optimum separation of the two groups.

Coefficients „*a*“ of discriminant function

unstandardized coefficients of disriminant function

not adjusted - *raw coefficients*

adjusted

The adjusted coefficients are modified so that the origin of the discriminant function (i.e., the point where all canonical axes have zero values) is located at the grand centroid, that is, at the point of the average values of all characters.

standardized coefficients

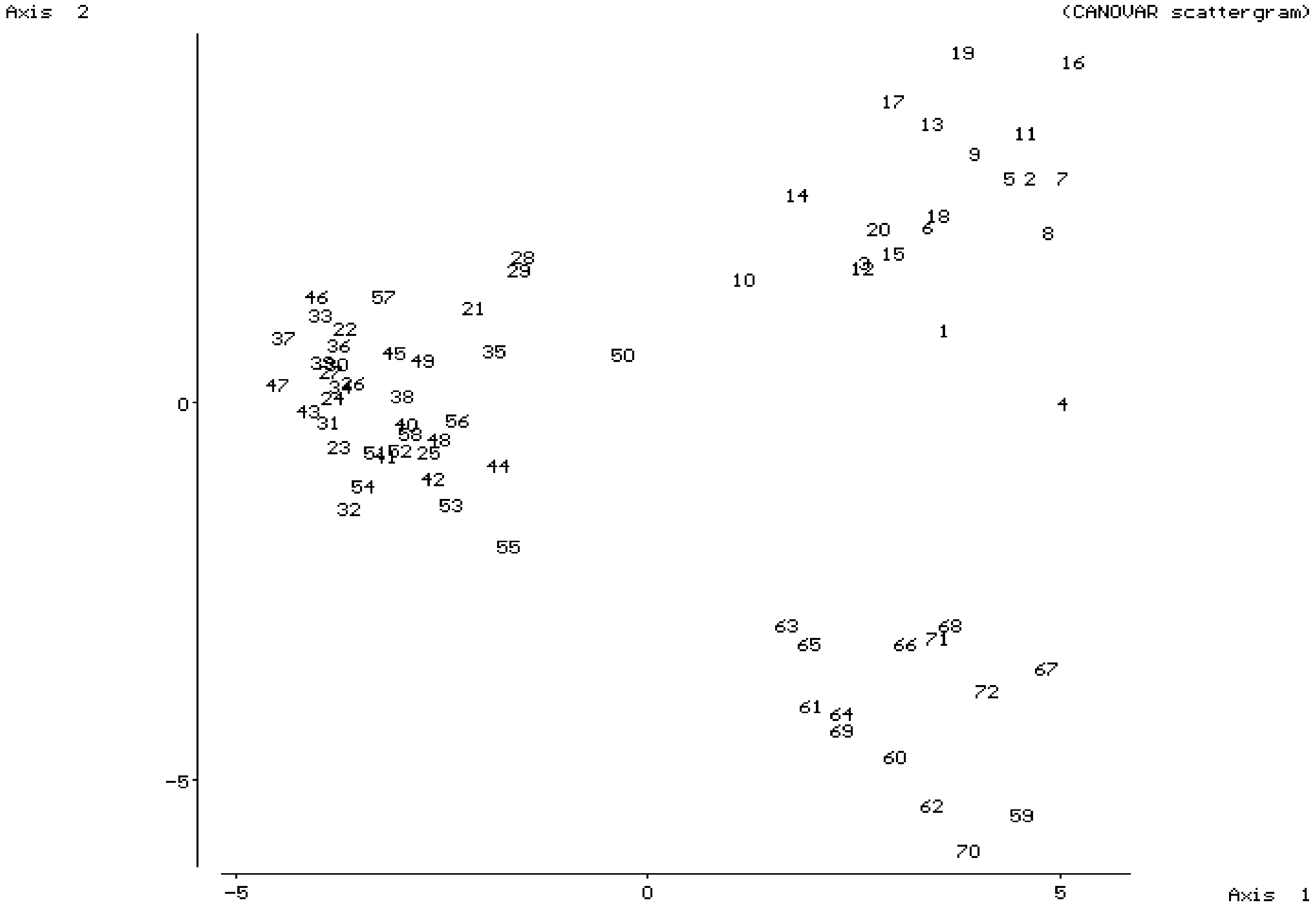
Maximum number of discriminant functions

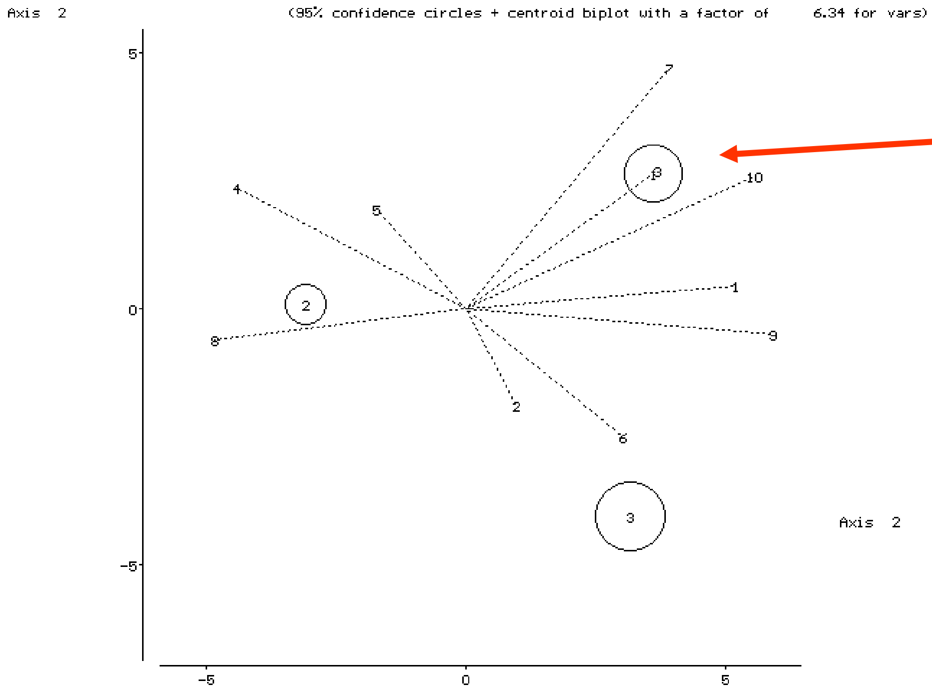
(maximum number of axes, maximum number of non-zero eigenvalues)

$s = \min(p; g - 1)$

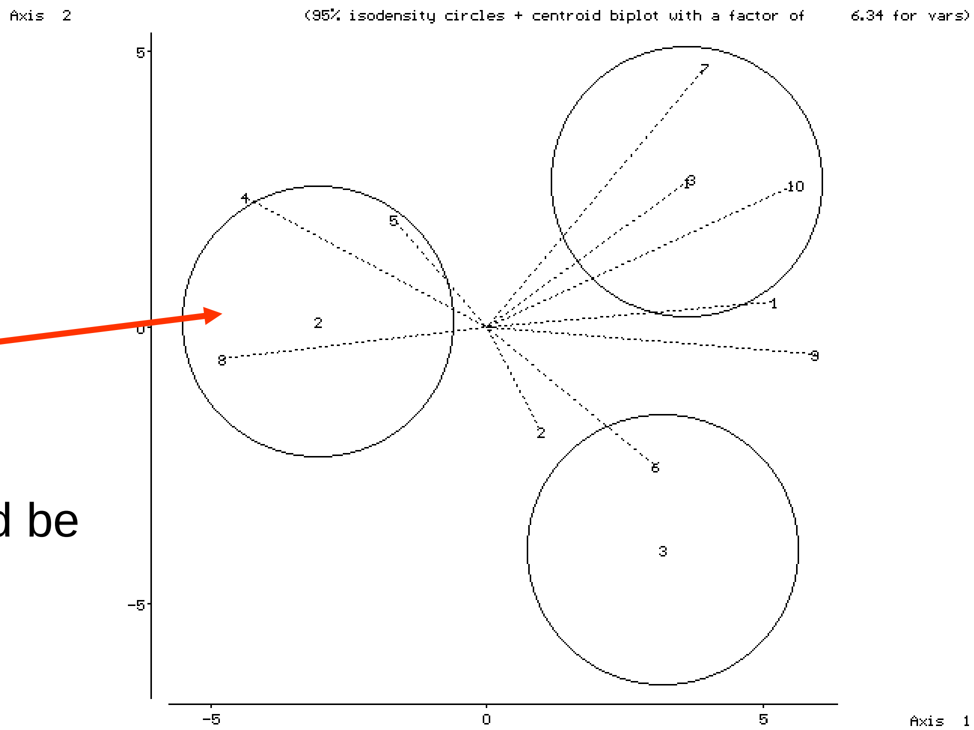
Interpretation of canonical axes (canonical discriminant functions)

(a) Relative position of objects, position of centroids





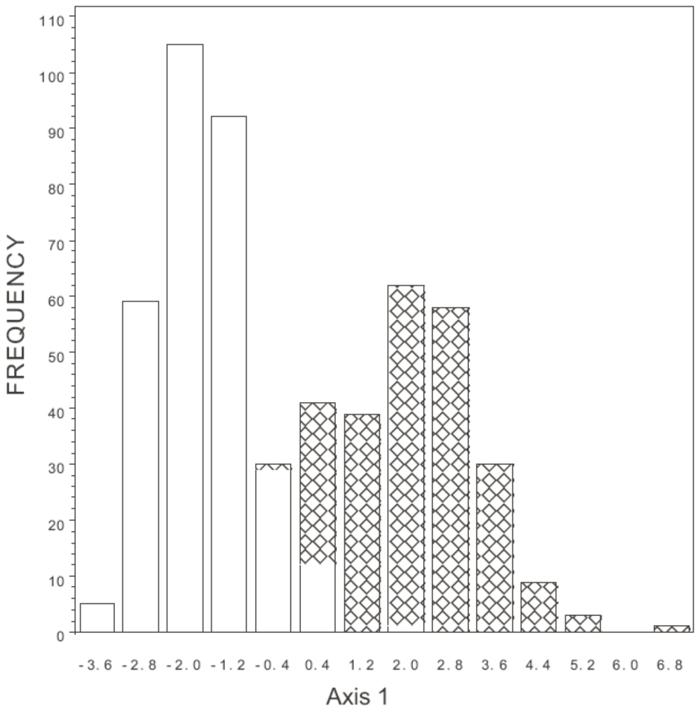
95 % Confidence interval of the centroid

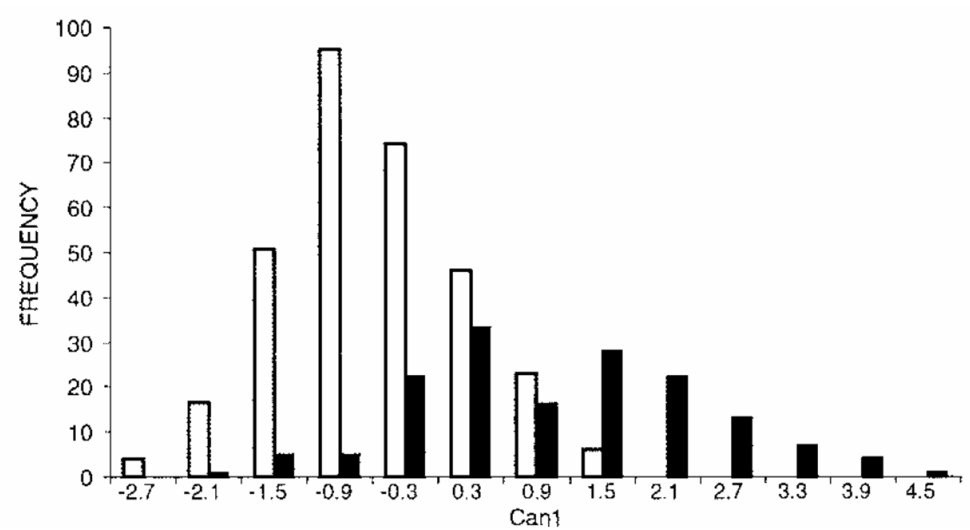


The area in which 95% of the objects of a given group should be located (assuming a normal distribution of characters).

Interpretation of canonical axes (canonical discriminant functions)

(a) Relative position of objects





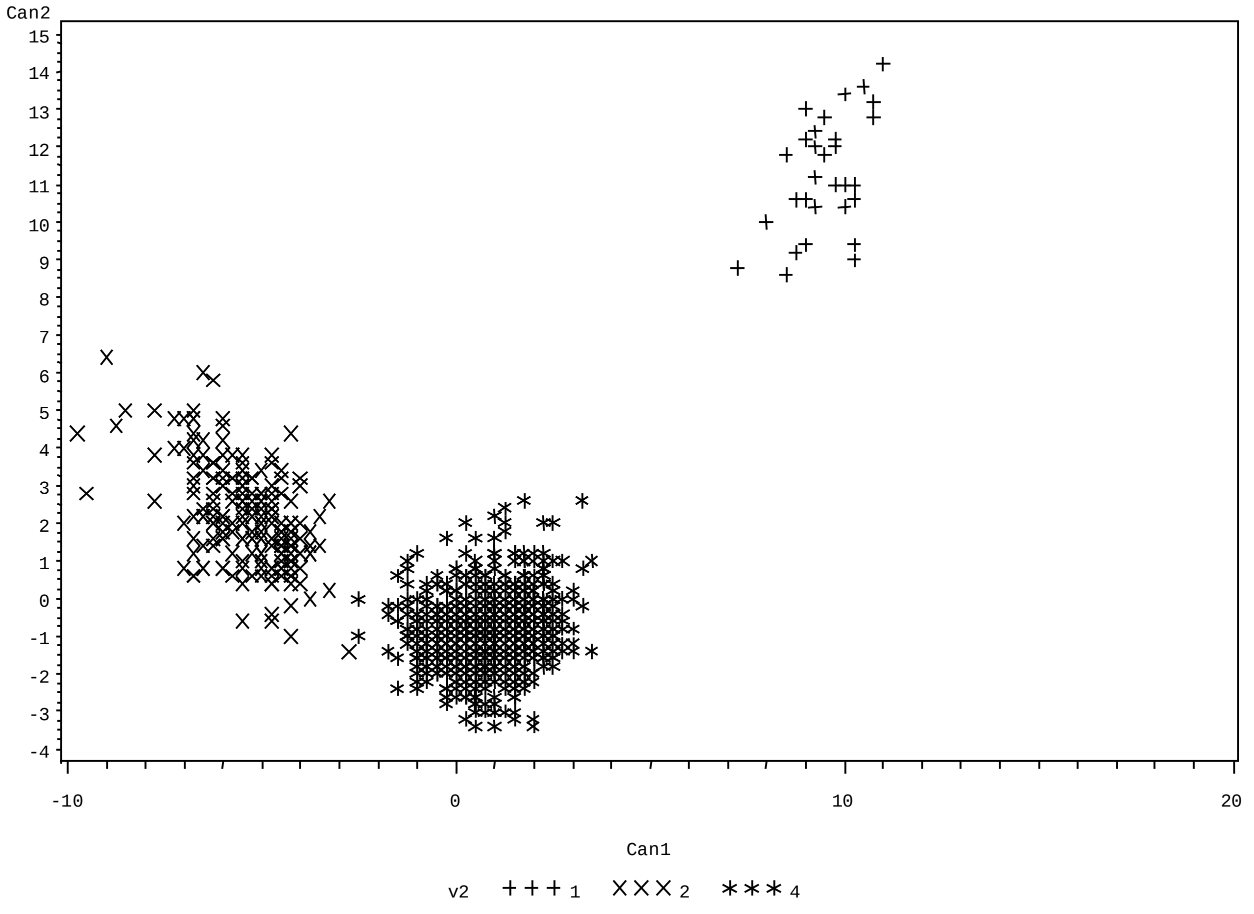
(b) total structure coefficients, total canonical structure

(c) eigenvalues

(d) canonical correlation coefficients

The square of the canonical correlation coefficients can be interpreted as the proportion of the variability of the discriminant function explained by the groups or the differences between the groups. This characteristic can sometimes be more useful than the percentage representation of eigenvalues. If the groups differ only slightly in the analyzed characteristics, the values of the canonical correlation coefficients will be low.

(e) The statistical significance of discriminant functions (axes) is evaluated using the criteria of Wilks‘ lambda, chi-square, or the likelihood ratio.



diploid *O. macrocarpon* +

diploid *O. microcarpus* x

polyploid *O. palustris* *

Oxycoccus - kanonicka diskriminacni analyza

The CANDISC Procedure

	Canonical correlation	Adjusted canonical correlation	Approximate standard error	Squared canonical correlation
1	0.942491	0.940682	0.003217	0.888290
2	0.905916	0.903497	0.005164	0.820683

Test of H0: The canonical correlations in the current row and all that follow are zero

Values of Inv(E)*H = CanRsq/(1-CanRsq)

	Eigenvalue	Difference	Proportion	Cumulative	Likelihood ratio	Approximate F Value	Num DF	Den DF	Pr > F
1	7.9518	3.3750	0.6347	0.6347	0.02003149	202.76	70	2340	<.0001
2	4.5767		0.3653	1.0000	.17931699	157.63	34	1171	<.0001

The CANDISC Procedure

Total Canonical Structure

Variable	Can1	Can2
v4	0.672599	0.282658
v5	0.712117	0.099797
v6	0.683018	0.232456
v7	0.693296	0.143058
v8	**0.814472**	0.054974
v9	0.542881	-0.217714
v10	0.266363	-0.256140
v11	0.361661	-0.180027
v12	**0.830531**	-0.094723
v13	-0.126190	-0.090211
v14	0.663007	0.090076
v15	0.760576	-0.095871
v16	0.593574	0.269126
v17	0.451203	-0.111711
v18	0.725532	**0.417065**
v19	0.361339	-0.197338
v20	0.116087	-0.283614
v21	0.512043	**0.345655**
v22	-0.045555	0.264878

(3) Total-Sample Standardized Canonical Coefficients

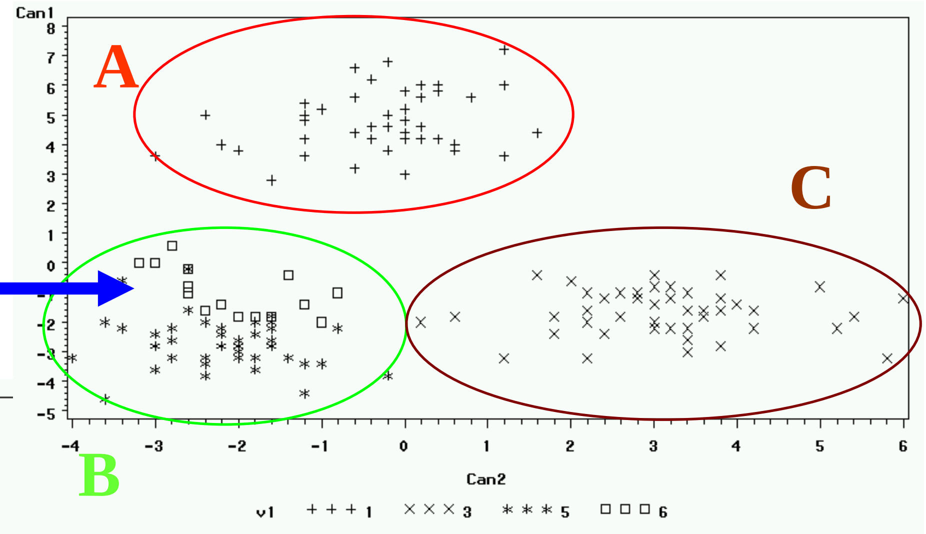
Variable	Can1	Can2
v4	2.20664064	1.14717014
v5	-1.29441256	**-7.28758830**
v6	-1.88656987	1.41224150
v7	1.19813550	**6.97436468**
v8	0.88095578	0.97932782
v9	-0.53521212	-1.25167526
v10	0.16348543	-0.20978626
v11	-0.48940387	-0.02370969
v12	0.34296483	-0.19078962
v13	0.21085257	-0.63991444
v14	0.20793931	0.40710833
v15	0.17302770	-0.24243622
v16	0.13295992	0.41005661
v17	-0.02748186	-0.00185147
v18	0.42290424	0.71065636
v19	**11.85793927**	**-5.40534569**
v20	**14.40526450**	**-6.42814071**
v21	0.19589232	0.29079279
v22	-0.27396644	-1.17566832

Oxycoccus - kanonicka diskriminacni analyza
The CANDISC Procedure

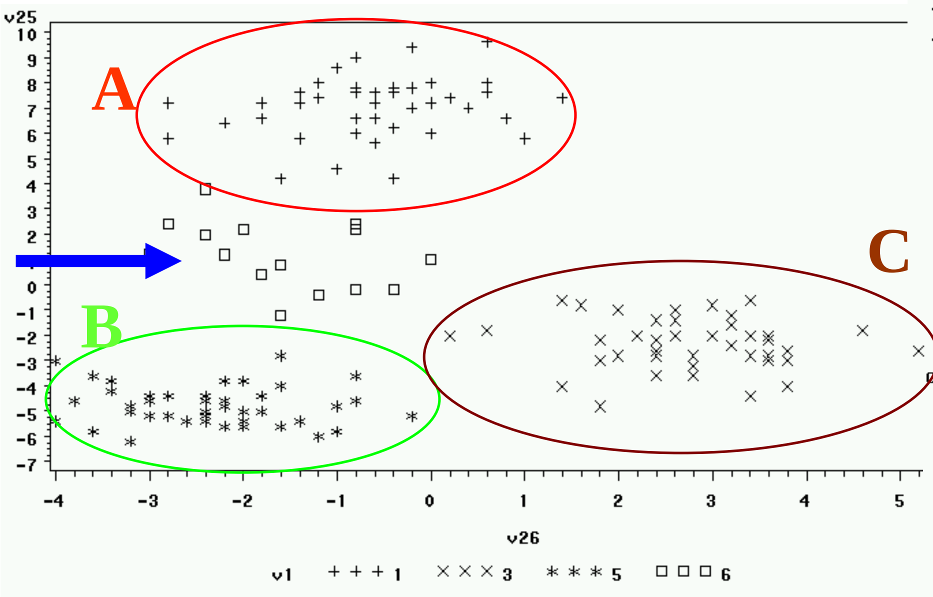
Raw Canonical Coefficients

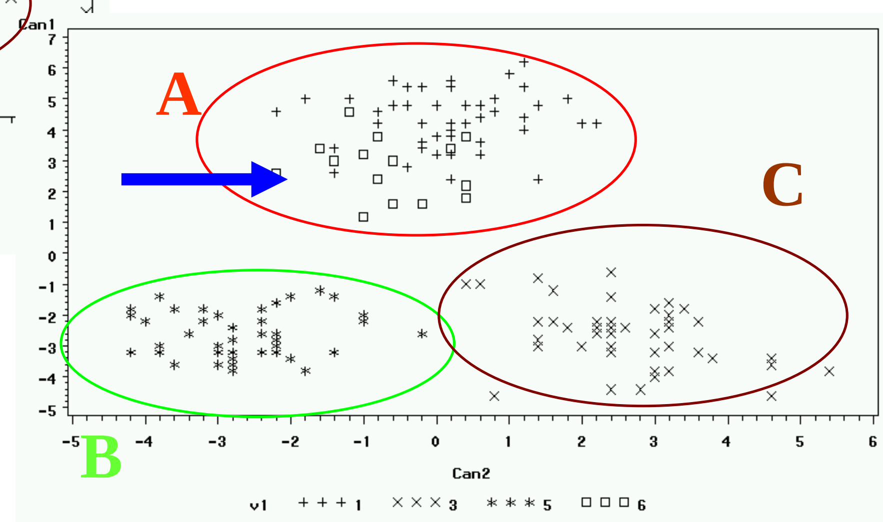
Variable	Can1	Can2
v4	0.838448541	0.435885713
v5	-1.304051753	-7.341857294
v6	-1.452234796	1.087108555
v7	1.134513637	6.604020879
v8	0.784902149	0.872548345
v9	-0.992564918	-2.321264617
V10	0.689107945	-0.884270705
v11	-2.269512407	-0.109948923
v12	0.454311114	-0.252730994
v13	0.793906003	-2.409417741
v14	0.525571082	1.028975078
v15	0.176677291	-0.247549821
v16	0.242091319	0.746624615
v17	-0.102516110	-0.006906569
v18	0.660204493	1.109420248
v19	1.877413657	-0.855803828
v20	1.939521377	-0.865483332
v21	0.199233118	0.295752053
v22	-0.581907220	-2.497130258

□ Not assigned to any group



□ Assigned to B





□ Assigned to A

Note: Assigning transitional objects to different groups may yield different discriminant functions and different results.

Classificatory discriminant analysis

(a) Searching for an identification (classification) criterion

Groups of objects with known classification

Group of objects with uncertain status

(b) Ascertaining the effectiveness of classification criterion

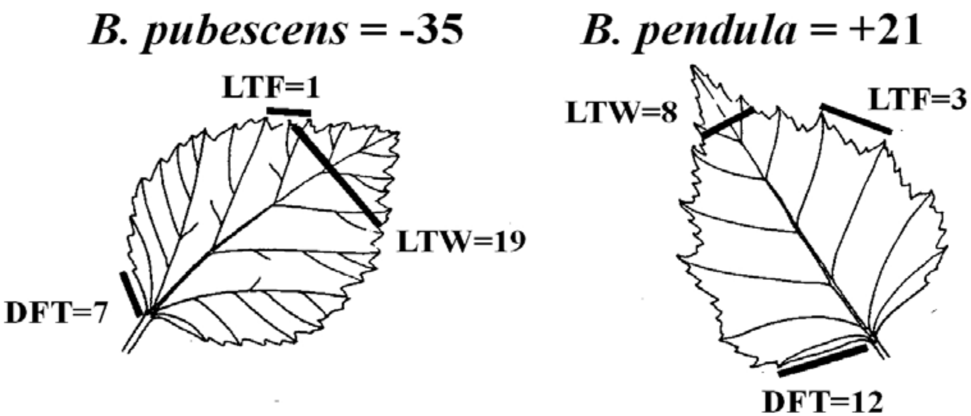
resubstitution

cross-validation

The effectiveness of the classification criterion is tested on the same dataset from which this classification rule is derived (this testing method is called resubstitution). If we have a smaller number of objects, it is advisable to use cross-validation: From a dataset of n objects, we select $n−1$objects, which we use as the training set. Based on this training set, we derive the classification criterion, which we then apply to the one omitted case. We repeat this procedure n times.

Methods of deriving the classification rule:

(1) Canonical discriminant function - objects are classified based on their score on the canonical discriminant function or based on their projection into canonical space

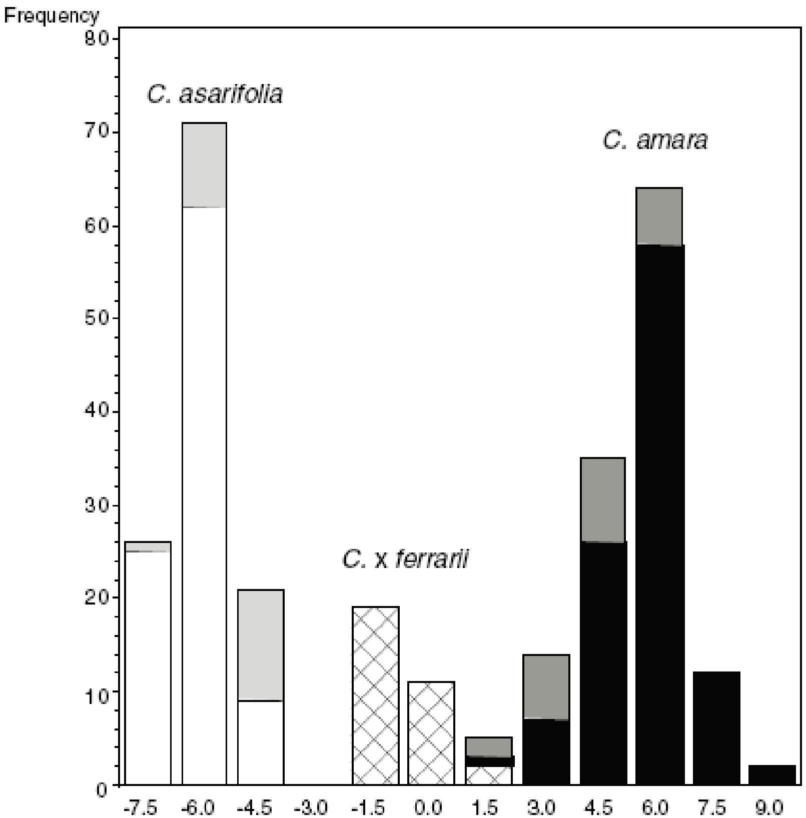


Discriminant function for the identification of *Betula pubescens* and *B. pendula*
12LTF + 2DFT – 2LTW – 23
Positive values *B. pendula*, negative values *B. pubescens*
Probability of correct identification 93%
(Stace, C. A., 1991, New Flora of the British Isles)

Methods of deriving the classification rule:

(1) Canonical discriminant function - objects are classified based on their score on the canonical discriminant function or based on their projection into canonical space

The classified object is displayed in canonical space along with a set of known objects (whose group membership is known). Based on the relative position of the classified object and the set of known objects, the membership of this element to a particular group is inferred.



(2) Calculation of the linear classification function for each group

A separate linear classification function is calculated for each group of objects. The classification score of the unknown (classified) object is then calculated for each of these functions. The object will be assigned to the group for which the classification score reaches the highest value.

(3) Classification rules based on probabilistic models

(i) Linear discriminant function
(ii) Quadratic discriminant function
(iii) Nonparametric methods, e.g., *k*-nearest neighbors

(3) Classification rules based on probabilistic models

- ***v*** is generally a *p*-component vector of characters
- $\boldsymbol{v}_0$ represents a vector of characters of a specific object
- ***v*** has different probabilities of belonging to π_1 and π_2
- probability densities - $f_1(\boldsymbol{v})$ for π_1 and $f_2(\boldsymbol{v})$ for π_2.
- the space R containing all objects has subspaces R_1 and R_2,
  it holds that $R_1 \cap R_2 = 0$ and $R = R_1 \cup R_2$)
- a classification rule will define the division of space R into two mutually exclusive subspaces R_1 and R_2, and at the same time will assign objects from group π_1 to R_1 and objects from group π_2 to R_2.

Subspace R_1 is defined as the set of vectors $\boldsymbol{v}$,

for which: $f_1(\boldsymbol{v}) > f_2(\boldsymbol{v})$

Subspace R_2 is defined as the set of vectors $\boldsymbol{v}$,

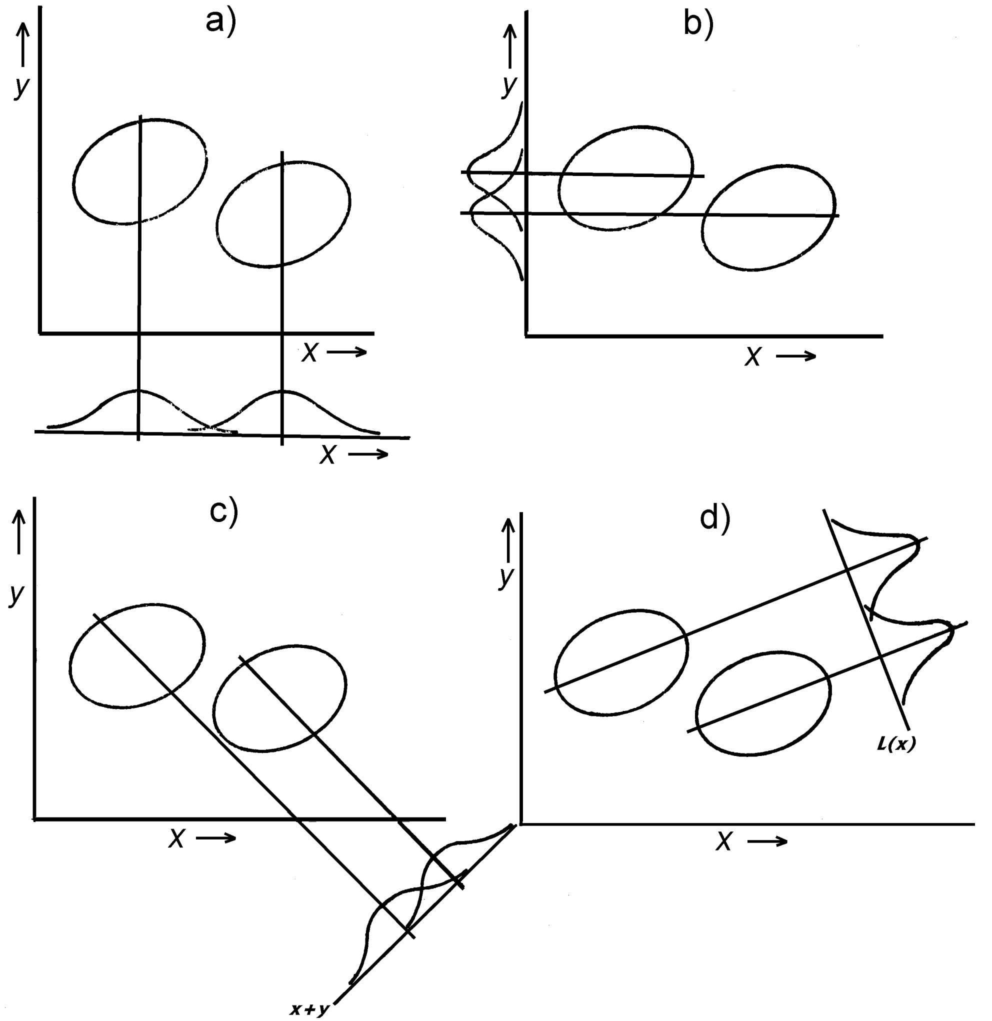
for which: $f_1(\boldsymbol{v}) \leq f_2(\boldsymbol{v})$.

The values of $f_1(\boldsymbol{v})$ and $f_2(\boldsymbol{v})$ can be estimated based on the results of training set measurements.

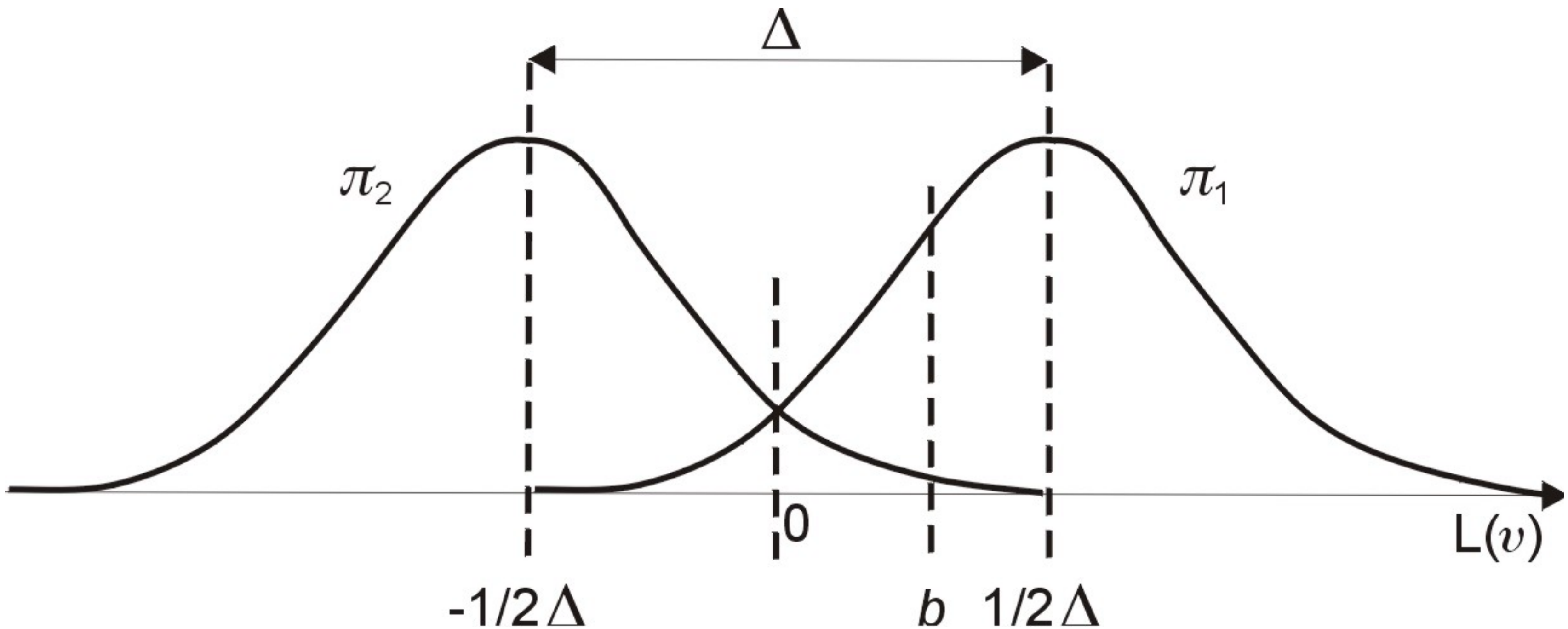
The classification rule then takes the form:

$\boldsymbol{v}$ belonds to π_1, if $f_1(\boldsymbol{v})/f_2(\boldsymbol{v}) > 1$

$\boldsymbol{v}$ belonds to π_2, if $f_1(\boldsymbol{v})/f_2(\boldsymbol{v}) \leq 1$



In these cases, $f_i(\boldsymbol{v})$ is based on the assumption that most objects are clustered around the center of mass (centroid), and their density decreases with increasing distance from the centroid.



L(v) – linear discriminant function

Δ – Mahalanobis distance expressing the distinction between groups π_1 and π_2

Number of Observations and Percent Classified into v2

From v2	1	2	4	Total
1	30	0	0	30
	100.00	0.00	0.00	100.00
2	0	209	1	210
	0.00	**99.52**	0.48	100.00
4	0	2	965	967
	0.00	0.21	**99.79**	100.00
Total	30	211	966	1207
	2.49	17.48	80.03	100.00
Priors	0.33333	0.33333	0.33333	

[1] Posterior Probability of Membership in v2

Obs	v2	From into v2		Classified 1	2	4
200	2	4	*	0.0000	0.4354	0.5646
437	4	2	*	0.0000	0.8598	0.1402
452	4	2	*	0.0000	0.5198	0.4802

* Misclassified observation

Stepwise discriminant analysis

Stepwise discriminant analysis seeks a combination of features that together allow the best possible separation of predefined groups.

The set of the most suitable features is selected gradually, in individual steps.

The method starts by selecting the feature that best separates the predefined groups; in the next step, it evaluates all remaining features and finds the one that best separates the groups in combination with the already selected feature.

At each step, the statistical significance of the selected features is calculated (the value 'F-to-remove,' statistics for removal) as well as the statistical significance of the remaining features (the value 'F-to-enter,' statistics for entry).