A NEW VARIANT OF DISCRIMINANT ANALYSIS AND ITS APPLICATION TO DISTINGUISHING FESTUCA POPULATIONS

by

FISCHER J.+ – HORÁNSZKY A.* – Mrs. KISS N.+ – SZÖCS Z.**

Received on 30th May, 1973

Introduction

When R. A. F i s h e r laid the foundations of discriminant analysis (F i s h e r 1936, 1938), he probably rather strove to solve a question of theory (optimum differentiation of multivariate populations of normal distribution) than to contribute a new method to taxonomy. Still, one could almost consider symbolic that he demonstrated the new procedure on the distinction between the populations of two *Iris* species *(I. setosa* and *I. versicolor)*, since DA also lends itself – among others – to solve one of the most difficult problems of taxonomy: that of differentiating between closely related taxons.

Although the use of DA gained ground in several domains of biology (e.g. in anthropology or zoosystematics) within relatively short time and meanwhile also the method itself was improved, plant taxonomy "discovered" its capabilities only twenty years later (C l i f f o r d – B i n e t 1954). After that also the instances of applying the method in plant taxonomy increased (M o r i s h i m a – O k a 1960, G a r d i n e r – J e f f e r s 1962, J e f f e r s – B l a c k 1964, J e f f e r s 1966, etc.). By now DA is one of the members of the "arsenal" of plant taxonomy esteemed all the world over.

With some delay, DA also found application in the domain of quantitative plant ecology (N o r r i s – B a r k h a m 1970).

In Hungary it was I. P r é c s é n y i who first called attention to DA (P r é c s é n y i 1960). His initiation found several followers (H o r á n s z k y 1960, S i m o n 1964, B o r h i d i – I s é p y 1966, S v á b 1969 and H o r á n s z k y – S z ő c s 1973).

+ Institute for Computer Science and Automation of the Hungarian Academy of Sciences, Budapest. *Department for Plant Taxonomy and Ecology of the Eötvös Loránd University, Budapest. **Research Institute for Botany of the Hungarian Academy of Sciences, Vácrátót

The methods described in these works differ in many respects. However, they have two common features:

1. each of them is founded on the original R. A. Fisher's model;
2. none of them is suitable for an adequate characterization of the "discriminating power" of the variables (or rather, in the authors' opinion: the methods offered for this are unsatisfactory).

The authors publish the present paper in the conviction that the version of DA evolved by J. Fischer, on the first application in plant taxonomy of which they report here, resembles the other versions of DA regarding the feature mentioned first, still in respect of the second feature it is more effectual than those. The version to be discussed here affords an opportunity to a manysided and reliable characterization of the "discriminating power" of the variables. Founded on this it becomes possible – among others – to determine an optimum group of variables by confronting the "expenses" (time, work, costs, etc.) with discriminating information to be obtained by them.

In this way – as compared with the versions of DA used thus far – the theoretical and practical efficiency of the examination can be increased to a significant degree which – in view of the fact that quantitative taxonomic examinations consume much work and time – is not a negligible advantage.

In the authors' opinion the method suggested here is therefore – yet also on account of its other properties – lending itself particularly well for a biosystematical application. Just for this reason they set forth here (maybe in greater detail than usual) the description of the method and program, as well as the way of their use (especially that of the interpretation) with the intention of making the taxonomical application of the method as easy as possible.

To be sure, the advantages and drawbacks of a method can be properly valued in the first place relying upon the experience gained in the course of its application. Therefore the authors would be thankful if those who try the method notified them about the experience gained and the opinion formed in this regard.

Although – as it also appears from what has been said so far – the present paper is basically of methodical character, still, the example serving the purpose of demonstration is more than a mere illustration.

The authors have applied DA for differentiating *Festuca* populations several times (Horánszky 1960, Horánszky–Szőcs 1973), consequently the present paper is at the same time an organic continuation of the studies on Festuca published up to the present. In both quoted instances different variants of DA were used (other than the present one) but the conditions of the examination and the considered variables were similar. In the second instance (just as in the present examination) the numerical part of the analysis was already conducted by means of a computer.

The whole of this work has been complied from the following contributions.

F i s c h e r: biometric modelling, mathematical theory and modifications, parts of programming and interpretation, writing the algorithm description, and the "Method and Results",

H o r á n s z k y: ideas and productory of data material, parts of interpretation, writing the "Materials",

K i s s: basic program, parts of running and modifying, explaining the program structure and use,

S z ő c s: ideas, parts of running, writing the "Introduction" and "Conclusions".

Materials

In the reserve "Prohibited Forest" near Ujszentmargita (Great Hungarian Plain) there were intensive production-biological examinations in course within the IBP through several years, on which a comprehensive work by Z ó l y o m i – M á t h é – P r é c s é n y i – S z ő c s (1972) and the publications quoted in same have rendered account.

According to the traditional professional opinion, the dominating grass-forming species of the alkali steppe meadow (Artemisio-Festucetum pseudovinae) and the neighbouring pasture of sheep (Achilleo-Festucetum pseudovinae) is the same species: *Festuca pseudovina* H a c k. ap. W i e s b. Still, – in the first place regarding the habit – the specimens in the meadow differ so much from those in the pasture that the question turns up automatically whether the species is the same in both sites.

For deciding the question the authors adopted, besides other kinds of examination, the new version of discriminant analysis as expounded in the present paper.

The authors collected 100 plants each from the two sites, yet these were not all suitable for a complete series of measurings. So the primary material of the examination was composed, eventually, by 80 specimens from pasture and 84 ones from meadows (populations "A" resp. "B").

Relying upon experience previously gained by them, the authors limited the examination to the morphological properties of the inflorescence. They took 12 properties into consideration, of which the first 3 referred to the panicle, the others to the second spike of those on the apex of the panicle, as counted from above ("localized sampling" – H o r á n s z k y 1970).

The examined properties (called "variables" in the next Chapter) were the following:

1. the length of the panicle (cm.),
2. the length of the longest side-branch on the lowest node of the panicle (cm.),
3. the length of the first internodium of the panicle (cm.),
4. the length of the outer glume (gluma inferior) (mm.),
5. the length of the inner glume (gluma superior) (mm.),
6. the length of the first flower of the spikelet (spicula) counted from below, that is: the length of the awn (palea inferior) (mm.),

7. the length of the second flower of the spikelet (spicula) counted from below, that is: the length of the awn (palea inferior) (mm.),
8. the length of the third flower of the spikelet (spicula) counted from below, that is: the length of the awn (palea inferior) (mm.),
9. the length of the arista belonging to the first awn (mm.),
10. the length of the arista belonging to the second awn (mm.),
11. the length of the arista belonging to the third awn (mm.),
12. the number of flowers in the spikelet.

The examined specimens were collected in the state of waxen ripeness, when the spikes were not yet disintegrated but the change in size of the parts of the inflorescence was already quite small.

Measurings were conducted under a Zeiss SM XX stereomicroscope, on panicles and/or spikes laid on millimetre paper, with properties 1–3 to an accuracy of 1 mm., with properties 4–11 to one of 0.1 mm. (cf. Csányi–Horánszky 1973).

Method and inference of calculation*

The determining power of the traits differentiating the plant groups exposed to the two different complexes of circumstances was examined as follows.

As a basic mathematical method for attaining a possibly good separation of populations the authors applied discriminant analysis. The family of computer programs to be described is actually much wider and apt for solving even more sophisticated problems, yet it is sufficient here to confine to its linear version. The reason for this is that linear discriminant analysis (LDA) has proved to fit well enough i.e., weighted combinations of the original characteristics give an adequate chance for optimum grouping.

In the sequel, not merely the production of the "optimally separating linear combination of properties" (OLC) known from the literature (e.g., C. Smith: Biomathematics, 1967) will be meant by performing LDA. Beyond the above LDA will be considered as containing several additional methods of evaluating goodness of discrimination. These latter aim at yielding statistical measures and tests to judge the success of discrimination in general and also for certain traits. Part of such calculation procedures applied here or to be used later, as well as their comprehension into a common program and the complex inference attached are the authors' own results.

The running of programs took place in two phases. In the first the OLC of all the 12 properties was established. The related statistical methods of the above character led to the design of several narrower groups of traits for the combinations of which the necessary LDA-s were performed

* A concise joint discussion of mathematical and computing aspects of the subject has been necessary, since new and unknown methods are also being dealt with in this paper.

within one program. It is worth mentioning that this family of programs is extended also to contain an optional principal component analysis. This latter version is to further the selection of property groups possibly coherent as regards the particular problem. In the present instance the results of the first program were unambiguons enough to omit that option.

The numbers standing for the examined traits correspond to the code given above. Thus the first program performed the LDA of properties 1 – 12, the second one those of

Table I.

1, 2, 3		6, 7		11, 12
1, 3		6, 7		11
1, 3		6, 7		
1, 2, 3				
1, 3				
		6, 7		11, 12
		6, 7		11
		6, 7		
	4, 5		8, 9, 10	

in the sequence of the rows.

Both programs evaluated here yield the following output items.

1. Trait averages
2. Analysis of variance by traits (ANOVA; between the two groups)
3. Frequency distributions of traits (with quarter-standard-deviation class width)
4. Discriminant coefficients (weight of the traits in OLC)
5. Frequency distribution of Z-values (individual OLC values multiplied to the order of magnitude of a thousand)
6. Multiplied Z-values in increasing order and their group averages
7. Prediction chances, discriminant threshold, percentages of hit and failure, significance of deviation from overlapping (between the two groups)
8. Analysis of variance for Z-values (between the two groups, with the value of the corresponding F test statistic)
9. Percentual decrease of D^2 (between-group distance measure) when omitting single traits
10. Decomposition of the D^2 generalized distance into effects of variables (in the context of single variables versus variable pairs)
11. Contributions of the variables to the D^2 generalized distance (by halving the above pair effects between the respective two variables)
12. Correlation matrix of group mean differences (from the correlation coefficients among average group distances)

Explanatory comments on the above aim at elucidating the principles and methods of evaluation. That primarily for this particular application and supposing the reader's simultaneous look at the Annex.

1. For both groups (A = pasture, B = meadow) the mean values of the traits figuring in that particular LDA are presented (and denoted by their original code numbers). (TABLE III., 1.) The differences of these means give a rough preliminary idea about the presumable discriminatory powers of the properties. A limit is given to the weights to be gained in the OLC as if the two group means are very near then the role of the trait in question cannot be important.

2. Variances (mean squares) between and within the groups together with the F-values being the ratios of the former are presented for each trait. Formulas:

$$s_b^2 = \frac{n_A n_B}{n_A + n_B} (\bar{x}_A - \bar{x}_B)^2 \tag{1a}$$

$$s_w^2 = \frac{\sum\limits_A (x-\bar{x})^2 + \sum\limits_B (x-\bar{x})^2}{N-2} \tag{1b}$$

$$F = \frac{s_b^2}{s_w^2} \tag{1c}$$

where $\bar{x}$ denotes the respective group averages and $N = n_A + n_B$ the number of all individuals (objects). (TABLE III., 2.). These results can be understood as discriminant analyses for single variables (traits). – From the order of magnitude of the F-values above one can uniquely infer the separating rank order of traits but deficiently as it is also influenced by their types of distribution and interrelationship.

3. Frequency distributions of the variables are printed. Intervals of width defined by the quarter of the standard deviation from all respective data are considered around the property means i.e.,

$$\bar{x} \pm k \frac{s}{4} \tag{2}$$

give the limits of the frequency classes (closed at the right). (TABLE IV., 3.). This is the basic version which makes possible an approximate assessment of the overlaps from the printing by groups. If necessary, the distributions over the two groups can be established and plotted in different ways of uniting adjacent classes (options for which see 2/b in the description of parameter cards). All the above can serve for deciding whether and how some of the variables should be substituted by their aptly specified functions i.e., what transformations bid prospects for refining the discrimination.

4. Looking for an OLC and performing a connected LDA is built into the programs in two basic versions. The "homoscedastic" version

means that the standard deviations and correlation coefficients of the variables or in other words the within-group covariances

$$(3a)\qquad cov_A(i,j) = \frac{\sum\limits_A (x_i - \bar{x}_i)(x_j - \bar{x}_j)}{n_A - 1}$$

and

$$(3b)\qquad cov_B(i,j) = \frac{\sum\limits_B (x_i - \bar{x}_i)(x_j - \bar{x}_j)}{n_B - 1}$$

are considered as estimates of the same theoretical value for all i, j in the two groups. Written in matrix form this means that the covariance matrices as two-dimensional schemes representing the whole of the statistical estimates in (3a) and (3b), respectively, i.e.,

$$(4a)\qquad A = \begin{pmatrix} cov_A(1,1) & \dots & cov_A(1,M) \\ \vdots & & \vdots \\ cov_A(M,1) & \dots & cov_A(M,M) \end{pmatrix}$$

and

$$(4b)\qquad B = \begin{pmatrix} cov_B(1,1) & \dots & cov_B(1,M) \\ \vdots & & \vdots \\ cov_B(M,1) & \dots & cov_B(M,M) \end{pmatrix}$$

are supposed to be empirically established approximations for the same theoretical covariance matrix. The "heteroscedastic" version is, in turn, based on the assumption that standard deviations and correlations are different in the two theoretical background populations. In the programs the headline code of the two versions is "/1" for the former and "/2" for the latter. A prior choice between the two hypotheses being rather uncertain, the authors thought necessary to create an optional possibility of corresponding parallel LDA procedures. This was done by a simultaneous elaboration of the mathematical basis to "/2". At present, however, the authors deem sufficient to restrict themselves to the type "/1" and to present the details only of that one. Thus the criterion of OLC is that from the linear combinations of trait values in question i.e, from the set

$$(5)\qquad \sum_i c_i x_i$$

it should select an optimum one. More explicitly this means that from the possible M-tuples of constant multipliers c_i a particular M-tuple has to be selected so that the "new variable" produced by (5) be of the maximum F-value in sense 2.)

This requirement of the between-group variance possibly many times surpassing the within-group one can be visualized as a demand of a relatively highest stretching apart of the two groups. The desired goal can be reached by calculating the so-called inverse covariance sum matrix

(6) $$C = (A+B)^{-1}$$

Denoting the elements of (6) by c_{ij}, the numbers

(7) $$\lambda_i = \sum_j [c_{ij}(\bar{x}_{Aj} - \bar{x}_{Bj})]$$

will figure as OLC coefficients for the traits coded with the respective i. These λ-values are called discriminant coefficients. (TABLE IV., 4.) The λ-s by themselves give merely a reckoning rule to the OLC and are practically irrelevant when discriminatory power is concerned as the scaling and spread of traits can be markedly different.

5. Z-values are defined as individual values of the discriminant scores

(8) $$Z_i = \sum_i \lambda_i x_i$$

with λ_i described in 4.). Hence, Z-denotes the *OLC* quantities for each object and in the case of a good separation Z-values have to be in a close connection with their quality of belonging to group A or B. The principle of printing the frequency distribution is similar to that sketched in 3.), yet the Z-values are presented in the form of integers (TABLE IV., 5.) the multipliers (divisors) necessary to which are given in the heading of program table 6. This integer form has been considered practical as the bulk of Z- values are positive with three valid decimals; so both accuracy and simplicity of reckoning with them are ensured. Apart from the (practically ineffective) roundings off the integer representation is completely justified which can be seen if one takes into account that constant times OLC is OLC itself as all averages and standard deviations are multiplied by the same constant.

6. The output is a presentation of several Z-values written in the integer form of 5.). Before the Z-values their respective identifiers may optionally figure (in our case actually stem numberings). For objects of A the left, for those of B the right Z-column is used and so are printed all the individual values in increasing sequence. Below the group means $\bar{Z}_A$ and $\bar{Z}_B$ are given. (TABLE V.). It should be mentioned here that with the inverse application of the coefficient in the heading which means reckoning with the original magnitudes of $\bar{Z}_A$ and $\bar{Z}_B$, one can get

(9) $$\bar{Z}_A - \bar{Z}_B = D^2$$

where D^2 "Mahalonobis' generalized distance" is a measure of discriminability proportional to the highest possible between/within group variance ratio, see (15). The original definition is

(10) $$D^2 = \sum_i \sum_j (\bar{x}_{Ai} - \bar{x}_{Bi})\, c_{ij}\, (\bar{x}_{Aj} - \bar{x}_{Bj})\,.$$

The equivalence to (9) can be easily derived from (7) and (8). From the table not only the depth of overlappings in the OLC-categorization for the two groups and the position of objects displaced in "erroneous" directions can be read off but if necessary also the illegally behaving individual values can be identified. This permits to draw conclusions by confronting with the description of the material as to how far "mathematics could be right" and perhaps to get ideas for a correction of the categorizing function.

7. (TABLE VI., 7.) Prediction chances are presented in the form of percentual frequencies corresponding to the "optimum separation". Their values are printed in the order of

$$(11a)\qquad \frac{100n\,(A \to A)}{n_A};\quad \frac{100n\,(A \to B)}{n_A}$$

$$(11b)\qquad \frac{100n\,(B \to A)}{n_B};\quad \frac{100n\,(B \to B)}{n_B}$$

The symbol n ( ) denoting the respective number of individual cases and $\to$ the fact of having been grouped from the (original) group before into that (mathematical) after the symbol. These chances of categorizing are determined on the basis of whether an individual value does or does not surpass the discriminant threshold; the relative frequencies of types of such cases have been formulated in (11). In the table the value of this threshold has an integer representation as explained in 5.). The choice of the discriminant threshold is determined by the requirement that the hit percentage

$$(12)\qquad P_h = \frac{\dfrac{100n\,(A \to A)}{n_A} + \dfrac{100n\,(B \to B)}{n_B}}{2} = P\,(\text{hit})\,\%$$

should be maximum i. e., the failure percentage 100-P_h taking on its minimum value. From P_h the expression

$$(13)\qquad p = \exp\left[-\frac{n_A\,n_B\,(2P_h - 100)^2}{5000N}\right]$$

can be calculated which (in consequence of theorems by Kolmogorov and Smirnov, see e.g. Rényi 1962) means the following probability. Suppose that the OLC is equally distributed in the groups A and B; how probable is then the event that in spite of that the separation of the two groups takes place with at least the hit percentage actually found. This is called the significance (level) of deviation from overlapping. Its value may be expected to be unusually small for a significance level as an optimum separation even if insufficient generally represents a situation very far from the total coincidence. In the authors' practice, good discriminations used to give p-values below 10^{-13}, very good ones below 10^{-15} and excellent ones under 10^{-18}.

8. (TABLE VI., 8.). The ANOVA of Z-values is done from the original (not rounded) Z numbers. A not completely regular analysis of variance is in question here. The quantities

$$\frac{n_A n_B}{N}(\overline{Z}_A - \overline{Z}_B)^2 = \frac{n_A n_B}{N} D^4 \tag{14a}$$

and

$$\sum_A (Z - \overline{Z})^2 + \sum_B (Z - \overline{Z})^2 = D^2 \tag{14b}$$

are between-and within-group sums of squares in the usual ANOVA sense indeed yet the degrees of freedom are not proper ones. The reason for this is that "mean squares" obtained when dividing (14a) and (14b) by M respectively $N-M-1$ produce a ratio which is F-distributed with these degrees of freedom. It is, however, necessary to notice that such F-values reflect a correct reliability order in the maximum between/within-group variance ratio sense theoretically only if the joint distribution of the initial traits and so the distribution of Z-values is Gaussian normal. This is, in fact, a prerequisite also for the complete exactitude of the whole LDA. Thus one may state that D^2, and the connected F-value

$$F = \frac{n_A n_B (N - M - 1)}{MN} D^2 \tag{15}$$

predict the potential intensity of separation basing on the whole of both distributions but with a lower reliability as far as overlapping tails are concerned. If this critical remark is taken into account then the rank order of discriminatory powers may usefully be evaluated on the basis of significance levels read off from the M and $N-M-1$ degree of freedom critical values in the F-table. It did not seam necessary to actually calculate those levels in the program as the reader will see this in the part on inference. – Confronting the p-value defined in (13) with the above, that one represents an order of significances which is not sensitive with respect to the type of distribution as a whole. This originates from p being based on the mere fact of registering how high a deviation of group distribution functions can be attained. The value of p is shaped most immediately by the overlapping part of the two distributions proper but how those are is more subject to chance because of the smaller number of cases. From all what has been told and in accordance with the authors' experience the following assertion is logical: as in the phases of research and data collection one can have practically no idea regarding validity and consequences of applying the above confronted viewpoints, one is doing his best with a combined evaluation of the tests represented by (13) and (15). For this aspect, the reader is referred to the text commenting on (21) and (22).

9. The values of D^2 are calculated for the whole series of cases in which one of the variables is excluded from OLC determination. The modified D^2 quantities corresponding to the OLC of the $M-1$ remaining traits of choice are all less than the original M-variate D^2. As a matter of fact, by ommitting the k-th variable of the M

$$D_k^2 = D^2 - \frac{\lambda_k^2}{c_{kk}} \tag{16a}$$

is gained, c_{kk} denoting the respective diagonal element of the inverse covariance sum matrix C. The quantities

$$\delta D_k^2\,\% = \frac{100\,\lambda_k^2}{c_{kk}} \tag{16b}$$

are called the percentual decreases of D^2 and printed after the original code numbers of the traits denoted with $k = 1, \ldots . M$ in the actual configuration. (TABLE VI., 9.). The increasing order of δD_k^2-% points to a growing importance of variables. This in the sense how informative it is to introduce the respective trait into separation in addition to the others. Under the assumptions of LDA the above mentioned increasing order does exactly coincide with that of discrimination information gain as $\delta D_k^2\%$ is a monotonous function of the latter (defined in Kullback 1962). This very property exempts $\delta D_k^2\%$ as most usefully contributing to the complex LDA evaluation. It does not import immediate information about groups of variables still in the qualification it gives to single traits the connections with the others are regarded. For this reason the authors have introduced it.

10. Different scaling and dispersion of discriminant coefficients (which have been mentioned in 4.) can be taken into account and balanced for. This is done by modified values of which the formula is

$$B_i = \lambda_i \sqrt{\sum_A (x-\bar{x})^2 + \sum_B (x-\bar{x})^2} \tag{17}$$

These are the so called standard discriminant coefficients (or: standard regression coefficients). The following identity holding for them,

$$D^2 = \sum_i \sum_j (B_i B_j r_{ij}) \tag{18}$$

with the within-group correlation coefficients taken as r_{ij} are used in the table under the title "Decomposition of the D^2 generalized distance into contributions of variables". The decomposition is performed here in the sense of

$$D^2 = \sum_i B_i^2 + \sum_{j>i} 2B_i B_j r_{ij} \tag{19}$$

and the components are printed in a corresponding practical sequence. (TABLE VII.). The quantities $B_i{}^2$ may be interpreted as effects of single variables with $2B_iB_jr_{ij}$ variable-pair effects as their contrasts. The depth of decomposition i.e., that pairs are still concerned but more complicated sets are not is simple enough and gives an insight into inner connections already. This partition may produce negative components, too which contributes to the inference as to which variables can hinder each other in the discrimination. Small items suggest week contributions, large numbers point to stronger ones. Thus, following the decreasing order of components, a good implementation of information gained from the results in 9.) is established. This can be done either from the "absolute" values mentioned above or from the "relative" (to D^2) ones. At the end of these two columns figure the sums D^2 and 1, respectively.

11. The effects of variables by themselves and of pairs of them as treated in 10.) can be contracted to components of D^2 with the interpretation "total trait effect" for each variable. Let

$$B_i^2 + \sum_{j \neq i} B_i B_j r_{ij} \tag{20}$$

= "direct" + "indirect" effect be the characteristic for the i-th property. This is equivalent to summing up the i-th single-variable effect of 10.) as "direct" and all the halves of pair effects with x_i as one variable in the pair under the name "indirect" effect. For the whole system this means that the paired effects in 10.) are equally distributed between the two variables of that pair. It is obvious that in this decomposition the total sum is also D^2 and that negative components may occur. In the table, total effects are presented in the form of sums of the respective direct and indirect effects and all three kinds of values are expressed as percentages of the D^2 (TABLE VIII., 11.). In comparison to 10.) the table contains new aspects by making possible to draw one or more circles or at least relations of traits according to which those have more or less positive or negative effect on the discriminatory power of each other.

12. It has been mentioned under 4.) that the basis of LDA is an adapt combination of traitwise mean differences $\bar{x}_{Ai} - \bar{x}_{Bi}$ of a kind that a possibly high combined average group difference is produced in the form of the OLC-generated quantity $D^2 = \bar{Z}_A - \bar{Z}_B$. In accordance with this when the interdependence of the between-group differences $\bar{x}_{Ai} - \bar{x}_{Bi}$ is being characterized then this actually means the detection of their potentially interactive roles in attaining the actual D^2. The correlation coefficients of the mean differences which are to describe this very connection can be estimated even on the basis of one observation series just as the standard errors of means. In our case called homoscedastic in 4.) these coefficients are equal to the r_{ij} treated in 10.) (TABLE VIII., 12.). The heading of the Table refers to the heteroscedastic case where the way of computation is different and to the inference based on relations from the discriminatory point of view. The absolute value of the correlation coefficients (ranging as such from -1 to 1) is a measure of the pairwise connections they have to describe. The word measure is meant here in the sense of narrowness of

discriminatory connection; the sign of the coefficients corresponds to whether the roles of variables are parallel or antiparallel. It is natural that the reality of the above in details and as a whole depends on how far the assumptions contained in 8.) are fulfilled. Anyway, there can develop groups (sets) of variables within which correlations are high as referred to those connecting the variables of that set with variables not included in it. Trait groups shaped that way can be considered discriminating factors. The Table already given shows and makes use of this fact for our particular example. It is worth mentioning that this correlation matrix can serve for the design of less within-set correlated variables as representatives for the factors further in less trivial situations to optimally generate uncorrelated linear combinations of traits with the method of principal component analysis.

Summarizing the consequences drawn from all what has been outlined in the above 12 paragraphs and applied to the present example, the following can be stated. The evaluation by the program of sets of traits specified on the basis of the results yielded by the first program gave a practically unambiguous valuation of all the traits concerned. This situation prevailed already on the basis of statistical test quantities (P_h and F) with a simultaneous consideration of the number of properties to be measured (M) and taking their registration costs near to each other.

In the table following below the values P_h and F pertaining to the OLC-s of several trait sets examined are presented. As an information for the reader, let be mentioned that

$$F^* = (F-1)\sqrt{M} \tag{21}$$

is a good approximate comparative statistic of F-values in a sense that an F for which F^* is much greater than for the F of another trait group may be considered a better value. These values as well as some measures of information are computed and printed in more sophisticated versions of our programs. Most harmonizing with our present way of inference among those is

$$\frac{I(P_h)+I\left[\Phi\left(\frac{D}{2}\sqrt{N-M-1}\right)\right]}{2\sqrt{M}} \tag{22}$$

I denoting Shannon's information and Φ the standard normal distribution function. Quantities of a similar type help to decide in more complicated examples than ours which traits are worth to be included in future repetitions of that particular discrimination trial.

The Table, relying on which in our case consequences of such character can be drawn is the following:

Table II.

No.	Trait group												M	P_h	F
1.	1	2	3	4	5	6	7	8	9	10	11	12	12	96,9	30,2
2.	1	2	3			6	7				11	12	7	95,7	52,9
3.	1		3			6	7				11		5	95,1	66,3
4.	1		3			6	7						4	95,7	82,3
5.	1	2	3										3	92,1	99,7
6.	1		3										2	92,1	146,1
7.						6	7				11	12	4	75,9	15,7
8.						6	7				11		3	73,5	16,9
9.						6	7						2	72,6	23,7
10.				4	5			8	9	10			5	68,2	6,1

The comparison of the first and the last line shows that properties 4, 5, 8, 9, 10 are superfluous for answering our question. A survey of the lines 2. to 4. "sifts out" the traits 2, 11, 12 if one makes use of (21) and also takes the results of lines 5–6. into consideration. All these confronted with the forelast line prove that there is too little information contained in the trait pair of traits 6, 7. Thus, measuring the variables 1 and 3 is economical and sufficient for the problem treated here.

The meaning of this last statement is nothing else than an exemption of the two traits for the present isolated problem. Comparative studies of other differences in situations are planned, too. Those will be liable to underline the roles of variables different from 1 and 3. Thus a complex analysis may suggest the necessity of a larger set of properties in the future. However, this latter fact does not alter the momentaneous value of the present evaluation.

For the reader more interested in the details of program construction, a skeletal description of a form with rather general options is to follow here. The bulk is written in ALGOL but FORTRAN subroutines for the addition, multiplication and inversion of matrices are used, too.

Preceding the input of the data to be analysed a number of parameter cards is read to pick out from the optional versions the particular one needed to the actual task.

The parameter cards are:

1. Cards containing integer type data:

   MM the (initial) number of traits

   N number of all data (from both groups)

   $N1$ size of the first group

   LL = 1: the version treated here does not contain program parts which would ask for functional transformations without any separate "function card" (see later)

SC = 1 asks for an LDA based on homoscedasticity i.e., on the equality of standard deviations and correlation coefficients between groups
= 2 asks for the LDA-s according to both 1 and 3, in this sequence
= 3 heteroscedastic LDA, equalities of 1 are not assumed

AZ = 0 there is no identifier at the end of data cards
= 1 ends of data cards bear identifiers which will be printed to Z-values (individual values of OLC)

SK = 0: our version does not imply automatical omission of variables for sparing computing time

DS number of LDA-s for the same data set within one "second type" program

2. Cards containing one-dimensional arrays.

a) $Y[J]$ = 0 if the program has to consider the J-th trait continuous to count frequencies belonging to intervals
= 1 if the frequencies of the J-trait are attached to discrete values

b) $ST[J]$ is a step multiplicity array the elements of which are integer multipliers of a basic step to yield the desired step width. The basic step is a quarter of the pooled standard deviation for $Y[J] = 0$ and unity for $Y[J] = 1$. Thus, e.g. $ST[J] = Z$ means intervals of $s/2$ for continuous and jumps of 2 for discrete frequency counting.

c) $FN[J]$ contains a number of frequently occurring functional transformation possibilities which can be separately applied for the respective J-th variables. FN is a real array consisting for all J of a part of integer numbers $FV[J]$ and a vector of constants $CV[J]$. In fact, the first digit of FN is FV, the code of the transform type while the rest is CV, a free constant to it. The transforms automatically contained (beside two others of possible actual choice) in the real procedure are the following:

if $FV = 1$	then	$CV \,.\, X$
if $FV = 2$	then	$CV \,.\, X^2$
if $FV = 3$	then	$CV \,.\, \sqrt{X}$
if $FV = 4$	then	$CV \,.\, \log X$
if $FV = 5$	then	CV/X
if $FV = 6$	then	$CV \,.\, \exp(-X)$
if $FV = 7$	then	$2 \arcsin \sqrt{X+CV}$

where X stands for the original trait value. For example if FV [3] was -34.125 then the transformation coded by 2 i.e., $CV . X^2$ was applied to the data of trait 3 with $CV = -4.125$. Remarks: zero must not be given for CV; if traits have to be used in original then their codes should be 11.

d) $CO[J]$ is the real array which consists of the "costs" of inserting the respective J-th traits. On the basis of these costs are the information quantities (22), (23) and the like computed for an economically efficient choice of variables to be recorded. If one does not want to differentiate one can take all CO-values as unity.

3. The integer matrix array OMIT $[I, J]$

On the first parameter card it was fixed by the value of DS how many LDA computations were desired on the same data set. Now, the two-dimensional array OMIT consists of DS subsequent card images and is realised by cards telling for the I-th LDA which J-coded traits should figure in the analysis and which not.

OMIT $[I, J] = 0$ if from the I-th LDA the J-th variable is to be omitted;

OMIT $[I, J] = 1$ if in the I-th LDA the J-th variable is to be considered.

After the input of parameter cards the data cards themselves of the groups to be separated are read in. First the first and than immediately the second group gets into the memory. The end of the second one is marked by a number 333333. After this endmark of the task a card deck either of a new task on the same data or of a completely new task can follow from the first parameter card up to another marker with 333333 on it. A card with

99 1 1 1 1 1 1 1

signifies the end of the whole LDA series.

The program can be found ready for compilation and running on a disk. Its authors readily give all information needed for effective use. The running of the material presented here was performed on a CDC 3300 type computer in the Institute for Computer Science and Automation of the Hungarian Academy of Sciences.

Finally, to inform the reader about the computing time demand of our LDA programs: the following experience has been gained on a great many of group pairs with hundred-order sizes. The running of complete programs usually takes about one thousandth of a minute per date. Accordingly, the evaluation of the 12 traits of our 164 units by the first program lasted about 2 minutes while the analysis in the second program of altogether 35 variables run approximately 5.7 minutes.

Table III.

1–2. Discriminant analysis from 12 variables

1. Trait averages **2. Analysis of variance by traits**

	A	*B*	Source of variation	Total	Mean square and *F* values Between groups;	Within groups;	*F*
1.	44,723	28,394	1. MQ		10 925,4073	38,2856	285,37
2.	17,443	13,394	2. MQ		671,8061	6,8340	98,30
3.	11,490	6,781	3. MQ		908,7084	4,0429	224,76
4.	2,860	2,631	4. MQ		2,1463	0,1121	19,14
5.	1,964	1,884	5. MQ		0,2617	0,0718	3,64
6.	3,548	3,234	6. MQ		4,0206	0,0994	40,45
7.	3,497	3,200	7. MQ		3,6150	0,1133	31,89
8.	3,226	3,009	8. MQ		1,9262	0,0965	19,96
9.	0,693	0,651	9. MQ		0,0730	0,0664	1,10
10.	1,258	1,203	10. MQ		0,1249	0,1046	1,19
11.	1,568	1,397	11. MQ		1,2063	0,0919	13,13
12.	3,881	4,112	12. MQ		2,1969	0,7210	3,05
			FG	163	1	162	

Table IV.

3. Frequency distribution of trait 1

	A	*B*		*A*	*B*
11,81–	0	1	36,56–	3	3
13,35–	0	0	38,11–	5	0
14,90–	0	1	39,65–	5	0
16,45–	0	1	41,20–	7	2
18,00–	0	2	42,75–	8	1
19,54–	0	0	44,29–	7	0
21,09–	0	3	45,84–	8	0
22,64–	0	6	47,39–	4	0
24,18–	0	8	48,93–	6	0
25,73–	0	13	50,48–	7	0
27,28–	0	10	52,03–	2	0
28,82–	0	10	53,57–	3	0
30,37–	0	6	55,12–	2	0
31,92–	2	6	56,67–	3	0
33,46–	5	4	58,21–	0	0
35,01–	6	3	59,76–	1	0

4. Discriminant coefficients (lambda values)

1.	0,002566
2.	– 0,002453
3.	0,003464
4.	– 0,006347
5.	0,001460
6.	0,008970
7.	0,015043
8.	0,005289
9.	– 0,001721
10.	0,000447
11.	0,010932
12.	– 0,006356

5. Frequency distribution of multiplied Z-values

	A	B
83 – 91	0	1
92 – 100	0	4
101 – 109	0	2
110 – 117	0	9
118 – 126	0	13
127 – 135	0	15
136 – 144	0	20
145 – 152	1	9
153 – 161	6	6
162 – 170	9	0
171 – 179	10	1
180 – 187	14	0
188 – 196	11	0
197 – 205	18	0
206 – 214	3	0
215 – 222	4	0
223 – 231	5	0
232 – 240	2	0
241 – 249	0	0
250 – 257	1	0

Table V.

6. Multiplied Z-values in increasing order (multiplier = 1000)

Number	Z1	Z2	Number	Z1	Z2	Number	Z1	Z2	Number	Z1	Z2
111		85	94		134	17	161		51	191	
164		95	113		134	26	161		20	191	
146		97	125		135	69	162		28	193	
123		98	103		136	40	165		55	193	
133		98	128		136	38	165		43	195	
142		101	143		136	45	166		68	196	
105		106	153		136	62	166		61	196	
120		109	115		136	72	166		73	197	
159		112	137		137	34	167		31	197	
157		112	134		137	29	168		80	198	
117		114	89		137	54	168		9	198	
197		115	96		138	53	170		15	198	
147		115	156		139	109		172	49	198	
149		116	136		139	67	173		79	200	
148		117	150		141	39	173		78	200	
106		117	118		142	23	174		84	202	
145		118	100		142	74	175		57	202	
127		119	138		142	12	175		16	202	
158		119	90		142	6	176		21	203	

Number	Z1	Z2	Number	Z1	Z2	Number	Z1	Z2	Number	Z1	Z2
152		121	110		142	50	177		64	203	
163		121	104		143	35	178		32	203	
119		121	151		143	33	178		10	204	
88		123	86		144	14	180		11	205	
122		123	112		146	81	180		83	206	
161		125	108		147	46	181		48	209	
129		125	160		147	1	181		2	212	
135		125	124		147	47	181		30	216	
162		125	7	147		19	182		71	217	
98		126	140		148	70	182		41	217	
102		127	155		149	36	182		44	218	
116		127	99		150	60	182		4	223	
141		128	107		151	37	183		76	225	
91		128	42	156		66	184		27	226	
93		128	131		156	58	184		63	229	
114		130	92		158	52	185		5	230	
126		131	121		159	59	185		65	234	
85		131	95		159	18	188		3	236	
154		131	77	159		56	188		13	250	
144		131	8	160		75	188				
101		132	22	160		82	188			Z1	Z2
130		133	139		160	24	189		Group		
132		134	87		160	25	190		means	190	131

Table VI.

7. Maximum separation statistics

Prediction chances				Discriminant threshold	Hit and failure per cent		Significance of deviation from overlapping
2,4	97,6	96,3	3,8	159	96,9	3,1	$4{,}359 \cdot 10^{-32}$

8. Analysis of variance for Z-values

Source of variation	SQ	FG	MQ	F
Total		163		
Between groups	0,14102	12	0,0117521	30,25
Within groups	0,05867	151	0,0003885	

9. Percentual decrease of D^2 when omitting single traits

1.	23,83
2.	5,06
3.	4,35
4.	0,62
5.	0,03
6.	1,20
7.	3,38
8.	0,38
9.	0,03
10.	0,00
11.	1,67
12.	6,03

Table VII.

10. Decomposition of the D^2 generalized distance into effects of variables

Component	Coefficient of determination abs.	rel.	Component	Coefficient of determination abs.	rel.
B 1 square	0,0408	0,696	2B 4 B 5 R 4 5	–0,0002	–0,003
B 2 square	0,0067	0,114	2B 4 B 6 R 4 6	–0,0006	–0,010
B 3 square	0,0079	0,134	2B 4 B 7 R 4 7	–0,0011	–0,018
B 4 square	0,0007	0,012	2B 4 B 8 R 4 8	–0,0004	–0,006
B 5 square	0,0000	0,000	2B 4 B 9 R 4 9	0,0000	0,002
B 6 square	0,0013	0,022	2B 4 B10 R 410	0,0000	0,000
B 7 square	0,0042	0,071	2B 4 B11 R 411	–0,0007	–0,012
B 8 square	0,0004	0,007	2B 4 B12 R 412	0,0011	0,018
B 9 square	0,0000	0,000	2B 5 B 6 R 5 6	0,0000	0,002
B10 square	0,0000	0,000	2B 5 B 7 R 5 7	0,0001	0,002
B11 square	0,0018	0,030	2B 5 B 8 R 5 8	0,0000	0,000
B12 square	0,0047	0,080	2B 5 B 9 R 5 9	0,0000	0,000
2B 1 B 2 R 1 2	–0,0218	–0,372	2B 5 B10 R 510	0,0000	0,000
2B 1 B 3 R 1 3	0,0279	0,475	2B 5 B11 R 511	0,0001	0,002
2B 1 B 4 R 1 4	–0,0025	–0,043	2B 5 B12 R 512	–0,0002	–0,003
2B 1 B 5 R 1 5	0,0002	0,003	2B 6 B 7 R 6 7	0,0025	0,042
2B 1 B 6 R 1 6	0,0005	0,009	2B 6 B 8 R 6 8	0,0007	0,013
2B 1 B 7 R 1 7	–0,0006	–0,011	2B 6 B 9 R 6 9	0,0000	0,000
2B 1 B 8 R 1 8	0,0004	0,007	2B 6 B10 R 610	0,0000	0,000
2B 1 B 9 R 1 9	0,0000	0,000	2B 6 B11 R 611	0,0008	0,014
2B 1 B10 R 110	0,0000	0,000	2B 6 B12 R 612	–0,0008	–0,014
2B 1 B11 R 111	0,0000	0,000	2B 7 B 8 R 7 8	0,0016	0,028
2B 1 B12 R 112	–0,0020	–0,033	2B 7 B 9 R 7 9	0,0000	0,000
2B 2 B 3 R 2 3	–0,0099	–0,169	2B 7 B10 R 710	0,0000	0,000
2B 2 B 4 R 2 4	0,0009	0,015	2B 7 B11 R 711	0,0008	0,014
2B 2 B 5 R 2 5	–0,0001	–0,002	2B 7 B12 R 712	–0,0019	–0,032
2B 2 B 6 R 2 6	–0,0009	–0,015	2B 8 B 9 R 8 9	0,0000	0,000
2B 2 B 7 R 2 7	–0,0014	–0,023	2B 8 B10 R 810	0,0000	0,000
2B 2 B 8 R 2 8	–0,0007	–0,012	2B 8 B11 R 811	0,0004	0,006
2B 2 B 9 R 2 9	0,0000	0,000	2B 8 B12 R 812	–0,0011	–0,018
2B 2 B10 R 210	0,0000	0,000	2B 9 B10 R 910	0,0000	0,000
2B 2 B11 R 211	–0,0003	–0,005	2B 9 B11 R 911	–0,0002	–0,004
2B 2 B12 R 212	0,0012	0,020	2B 9 B12 R 912	–0,0000	–0,001
2B 3 B 4 R 3 4	–0,0009	–0,016	2B10 B11 R1011	0,0000	0,002
2B 3 B 5 R 3 5	0,0000	0,000	2B10 B12 R1012	0,0000	0,000
2B 3 B 6 R 3 6	0,0007	0,011	2B11 B12 R1112	–0,0008	–0,014
2B 3 B 7 R 3 7	0,0006	0,010			
2B 3 B 8 R 3 8	0,0003	0,006			
2B 3 B 9 R 3 9	0,0000	0,000	Total	0,0587	1,000
2B 3 B10 R 310	0,0000	0,000			
2B 3 B11 R 311	0,0000	0,000			
2B 3 B12 R 312	–0,0018	–0,031			

Table VIII.

11. Contributions of the variables to the D^2 generalized distance

Trait	Percentage B2/D2	+ SUM(BI × BJ × RIJ)	= H
1.	69,608	1,812	71,420
2.	11,356	−28,286	−16,931
3.	13,394	14,410	27,804
4.	1,248	− 3,724	− 2,476
5.	0,042	0,157	0,199
6.	2,209	2,581	4,790
7.	7,083	0,534	7,616
8.	0,746	1,209	1,955
9.	0,054	− 0,178	− 0,124
10.	0,006	0,036	0,042
11.	3,031	0,166	3,197
12.	8,042	− 5,534	2,509

12. Correlation matrix of group mean differences

1,000	0,662	0,778	0,229	0,096	0,035	−0,024	0,051	−0,009	0,008	0,002	0,071
0,662	1,000	0,686	0,203	0,180	0,151	0,130	0,213	−0,027	0,042	0,044	0,103
0,778	0,686	1,000	0,198	0,054	0,102	0,054	0,091	−0,045	−0,003	0,008	0,151
0,229	0,203	0,198	1,000	0,592	0,309	0,307	0,334	0,322	0,363	0,320	0,286
0,096	0,180	0,054	0,592	1,000	0,278	0,216	0,270	0,216	0,269	0,295	0,236
0,035	0,151	0,102	0,309	0,278	1,000	0,531	0,495	−0,062	0,209	0,277	0,170
−0,024	0,130	0,054	0,307	0,216	0,531	1,000	0,607	0,054	−0,030	0,146	0,214
0,051	0,213	0,091	0,334	0,270	0,495	0,607	1,000	0,036	0,164	0,207	0,367
−0,009	−0,027	−0,045	0,322	0,216	−0,062	0,054	0,036	1,000	0,541	0,488	−0,090
0,008	0,042	−0,003	0,363	0,269	0,209	−0,030	0,164	0,541	1,000	0,601	0,135
0,002	0,044	0,008	0,320	0,295	0,277	0,146	0,207	0,488	0,601	1,000	0,147
0,071	0,103	0,151	0,286	0,236	0,170	0,214	0,367	−0,090	0,135	0,147	1,000

Conclusions

1. Relying upon the results it can be stated that the two *Festuca pseudovina* populations examined ("A" and "B") are very well separable by the method of the authors. Misclassifications are highly infrequent (only 5 among 164 cases).

2. The groups of variables (1 – 2 – 3, 4 – 5, 6 – 7 – 8, 9 – 10 – 11, 12) determined by the analysis of the correlation matrix of the differences of the averages (see Appendix) exactly correspond to the grouping to be expected. It also appears (e.g. from the comparison of lines 7 and 10 of Table III. 1. with line 1) that the variables of the "localized sampling" (properties 4 – 12) together could contribute to the separation to a much lesser degree than did the group of the first three variables.

3. It is properties 1, 2 and 3 the measurements of length of the inflorescence that have a deciding part in the discrimination. Compared with them properties 6 and 7 contribute but to a quite slight measure to separation. The part of the other 7 properties is insignificant.

4. Finer analysis showed further that also from this group only two properties are really significant: the 1st land the 3rd ones (the length of the panicle and that of its first internodium). These two jointly call forth a discrimination of the same measure as do the first three jointly, and this is hardly less intense than the discrimination founded on all properties. It should be noted that property 3 can be partly substituted by property 2: 1 and 2 yield hardly weaker result than 1 and 3.

5. Therefore, in examinations of similar character (in case of *Festuca pseudovina)* instead of the 12 properties examined by the authors it is sufficient to consider two (1 and 3, or possibly 1 and 2).

6. Essentially, properties 1 and 3 depend on the length of the inflorescence. It is easy to understand that it is just these which are the most important ones as regards discrimination, since the differences of site come to be expressed in the stature of the whole plant and through this also in the length of the inflorescence. The site of population "B" is much more favourable (thicker surface soil, less alkalinization, better supply with water, there is neither mowing nor pasturing there) than that of population "A" (alkaline, lowly yielding pasture of sheep).

7. In case of *Festuca pseudovina* the length measurements significant as to discrimination have no taxonomic value; on the other hand the "taxonomically significant" properties proved insignificant as to discriminance. Therefore, founded on the present examination and presuming that the properties taken into account in the taxonomic key are in truth of high taxonomic value, the authors consider the two populations as belonging to one species also in the future.

8. In comparison with other related species possibly just the traits will markedly discriminate which now have but slightly separated the two populations, or have had no discriminating effect at all. On the basis of the present examinations no further particulars can be given in this respect now.

9. In serial tasks of discriminating – which are, however, of similar character – one may attain significant saving in "costs" by means of preliminary examinations. On the other hand, without preliminary examinations one might leave out of consideration some important variables helping discrimination or unnecessarily examine certain redundant variables.

10. With the applied discrimination procedure the authors strove to develop an efficient and economical strategy. The essence of this is to try to find the optimum balance between the surplus information gained by introducing further variables and the "costs" expended to the introduction of these (see Formula 22).

11. It should be noted that in Formula 22 the examination of all variables has been taken into consideration at equal "costs". If the pro-

portion of the "costs" expended to the single variables can be given – if only estimation-like – (in the form of "weights") then, modifying Formula 22 accordingly (see Formula 23), the optimization procedure will be more efficient:

$$\frac{I\,[P\,(\text{prop.})] + I\left[\Phi\left(\frac{D}{2}\sqrt{N-M-1}\right)\right]}{2\sqrt{\sum W_i}} \tag{23}$$

(where W_i stands for the "costs" of the single properties).

Summary

The separation of *Festuca pseudovina* populations from neighbouring but ecologically differing sites, of populations dissimilar also as regards their habit was examined founded on 12 morphological properties of the inflorescence. The authors' new method of discriminant analysis proved most suitable for this purpose.

The procedure affords an effective and economical strategy for a manysided analysis of the discriminating information residing in the variables which is ensured by the complex way of evaluation described in points 9 to 12 of the methodological part. By successive analyses performed in this way also the group of variables to be considered the optimum one can be selected on the basis of confronting the expenses ("costs") and gain in information.

As it appears from the theoretical considerations and from the experience collected in the course of the application of the method (from information which has been in part also published in the present paper) the adopted procedure seems promising for the solution of the critical problems of taxonomy.

REFERENCES

Borhidi, A. – Isépy, I. 1966. Statistical studies on *Potentilla* species. Acta Bot. Hung. **12:** 221–239.

Clifford, H. T. – Bunet, P. E. 1954. A quantitative study of a presumed hybrid swarm between *Eucalyptus elaeophora* and *E. goniocalyx*. Aust. J. Bot. **2:** 325–336.

Csányi-Kovács Cs., – Horánszky, A. 1973. Charakterisierung der *Festuca*-Populationen aufgrund der Merkmale der Rispe. Annal. Univ. Scient. Budapest, **15:** 65–72.

Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. Ann. Eugen. **7:** 179–188.

Fisher, R. A. 1938. The statistical utilization of multiple measurements. Ann. Eugen. **8:** 376–386.

Gardiner, A. – Jeffers, J. N. R. 1962. Analysis of the collective species *Betula alba* L. on the basis of leaf measurements. Silvae Gen. **11:** 156–161.

Horánszky, A. 1960. Statistical studies of *Festuca* species. Annal. Univ. Scient. Budapest, **3:** 225–227.

Horánszky, A. 1970. *Festuca*-tanulmányok II. (*Festuca* Studien II.) Bot. Közlem. **57:** 207–216.

Horánszky, A. – Szőcs, Z. 1973. Die Trennung von *Festuca*-Populationen mit Hilfe der Diskriminanzanalyse. Annal. Univ. Scient. Budapest, **15:** 73–79.

Jeffers, J. N. R. 1966. Principal component analysis in taxonomic research. Rapp. Uppsats. Instn. Skoglig. Mat. Statist. **6:** 138–159.

Jeffers, J. N. R. – Black, T. M. 1964. An analysis of variability of *Pinus contorta*. Forestry, **36:** 199–218.

Kullback, S, 1962. Information theory and statistics. New York: Wiley.

Morishima, H. – Oka, H. I. 1960. The pattern of interspecific variation in the genus *Oryza:* its quantitative representation by statistical methods. Evolution, **14:** 153–165.

Norris, J. M – Barkham, J. P. 1970. A comparison of some Cotswold beechwoods using multiple-discriminant analysis. J. Ecol. **58:** 603–619.

Précsényi, I. 1960. Discriminantia-analízis alkalmazása növényrendszertani problémák megoldásában. (The application of discriminance analysis in the solution of problems of plant taxonomy.) Agrártud. Egy. Tájékoztatója. Gödöllő, **1:** 197–211.

Rényi, A. 1962. Wahrscheinlichkeitsrechnung mit einem Anhang über Informationstheorie. Berlin: Deutscher Verlag der Wissenschaften.

Simon, T. 1964. Entdeckung und Zönologie der *Festuca dalmatica* (Hack.) Richt. in Ungarn und ihr statistischer Vergleich mit ssp. *pseudodalmatica* (Kraj) Soó. Annal. Univ. Scient. Budapest, **7:** 143–156.

Smith, C. A. B. 1967. Biomathematics. London: Griffin.

Sváb, J. 1969. Diszkriminancia-analízis alkalmazása kenderfajták (populációk) megkülönböztetésére kisparcellás kísérletekben. [The use of discriminance analysis for differentiation between species (populations) of hemp in smallplot experiments.] Rostnövények. **3:** 5–19.