Behavior Research Methods & Instrumentation
1974, Vol. 6 (4), 443-445

Program Abstracts/Algorithms

Classification by linear and quadratic discriminant scores

MARK C. FULCOMER
New York University, New York, New York 10003

and

PETER H. SCHÖNEMANN and GEORGE MOLNAR
Purdue University, West Lafayette, Indiana 47907

This program provides two alternative Bayes solutions to problems of classifying an individual into one of K mutually exclusive populations on the basis of measurements taken on p predictor variables. It is assumed that the individual *must* have come from one of the K populations and *must* be assigned to *one* of them.

Two simplifying assumptions are made. First, the p measurements are assumed to have a multivariate normal distribution in each of the populations. Secondly, all misclassification errors are considered equally costly.

The Bayes decision rule minimizes the total probability of misclassification. In this procedure, an individual is classified by means of "discriminant scores," one for each of the K populations, resulting in the assignment of the individual to that population for which he has the largest posterior probability. Such a Bayes procedure requires the "prior probabilities" that an individual, drawn at random, belongs to a given population. This procedure does not, however, require that the covariance matrices of the K populations be equal (homogeneous); a test of the homogeneity assumption is made by the program. If they are *equal*, the discriminant scores can be reduced to linear functions of the predictor variables. They are therefore called "linear discriminant scores." When the covariance matrices are *unequal*, the discriminant scores are quadratic functions and are called "quadratic discriminant scores." Thus, the mathematical form of the discriminant scores differentiates two types of Bayes procedures—linear and quadratic—both of which are provided by the program. For detailed discussions, see Anderson (1958), Rao (1965), and Fulcomer (1970).

Method. (1) *Notation.* The following terms are used in this program description: K = number of populations, p = number of variables, $x' = (x_1, \cdots, x_p)$ = vector of predictor scores, π_k = prior probability of the k^{th} population, μ_k = mean vector of the k^{th} population, Σ_k = covariance matrix of the k^{th} population, r_{ij} = cost in assigning an individual who actually belongs to the i^{th} population to the j^{th}, $p_i(x)$ = probability density at x for the i^{th} population, § = sample space of all potential observations, ω_i = classification region for the i^{th} population. Carets (e.g., $\hat{\mu}_k$) indicate the use of sample estimates for corresponding parameters.

(2) *Decision Rule.* The expected loss in applying a decision rule for an individual from the i^{th} population is

$$R_i(x) = \sum_{j=1}^{K} \int_{\omega_j} r_{ij} p_i(x) dx.$$

The total expected loss

$$R(x) = \sum_{i=1}^{K} \pi_i R_i(x)$$

becomes

$$R(x) = \sum_{i=1}^{K} \int_{\omega_i} - S_i(x) dx$$

where

$$S_i(x) = -\sum_{j=1}^{K} \pi_j r_{ji} p_j(x)$$

is the i^{th} "discriminant score" of an individual with measurements x. Rao (1965) shows that if the classification regions ω_i of § are chosen such that $x \epsilon \omega_i \rightarrow S_i(x) \geqslant S_j(x)$ for all j, the total expected loss is minimized. Therefore, an optimal solution is one which places an individual into that population for which his discriminant score $S_i(x)$ is largest.

For a symmetric loss function (as used in this program), $r_{ij} = 1$ for $i \neq j$ and $r_{ii} = 0$, the discriminant score for the i^{th} population becomes

$$S_i(x) = -\sum_{j=1}^{K} \pi_j r_{ji} p_j(x) = -\sum_{j=1}^{K} \pi_j p_j(x) + \pi_i p_i(x)$$

$$= \pi_i p_i(x) + c = S_i^*(x) + c,$$

where c is a constant independent of i. The redefined discriminant scores $S_i^*(x)$ are used for classification. Since the posterior probability of the i^{th} population is

$$\pi_i p_i(x) \Big/ \sum_{i=1}^{K} \pi_i p_i(x) = S_i^*(x) \Big/ \sum_{i=1}^{K} \pi_i p_i(x)$$

and the denominator of this expression is a constant, an individual's largest discriminant score corresponds to that population for which he has the largest posterior probability.

(3) *Multivariate Normal Distributions.* If the distribution of x is p-variate multivariate normal in each

of the K populations with mean vector μ_k and covariance matrix Σ_k for $k = 1, \cdots, K$, the density at x in the i^{th} population is

$$p_i(x) = (2\pi)^{-p/2} \mid \Sigma_i \mid^{-1/2} \exp[-1/2(x-\mu_i)'\Sigma_i^{-1}(x-\mu_i)].$$

Although the discriminant score is $\pi_i p_i(x)$, it is more convenient to use the natural logarithm (ln) of the density and to omit the common factor $(2\pi)^{-p/2}$. The classification decisions are invariant under such a monotonic transformation, since it preserves the order relationships among the discriminant scores. Consequently, the i^{th} discriminant score may be redefined as

$$Q_i(x) = -1/2[\ln \mid \Sigma_i \mid] - 1/2(x-\mu_i)'\Sigma_i^{-1}(x-\mu_i) + \ln(\pi_i).$$

Since this score is a quadratic function of x, it is called a "quadratic discriminant score"; an individual is assigned to that population for which his quadratic discriminant score is largest.

(4) *Equal Covariance Matrices.* If the covariance matrices are homogeneous ($\Sigma_1 = \cdots = \Sigma_K = \Sigma$), further simplification of the discriminant scores results. The terms $-1/2[\ln \mid \Sigma_i \mid]$ and $-1/2[x'\Sigma_i^{-1}x]$ are common to all of the $Q_i(x)$ and may be omitted. Equivalent discriminant scores are

$$L_i(x) = -1/2(\mu_i'\Sigma^{-1}\mu_i) + \mu_i'\Sigma^{-1}x + \ln(\pi_i) \quad i = 1, \cdots, K$$

which are linear functions of the components of x and are, therefore, called "linear discriminant scores." $L_i(x)$ is analogous to the linear discriminant function originally derived by Fisher (1936). Although $L_i(x)$ is more familiar, it is apparent from the above developments that $Q_i(x)$ is more general since it is unlikely that the covariance matrices would ever be exactly homogeneous; one would expect quadratic discriminant scores, in general, to result in more accurate classification since they take such covariance matrix differences into account.

(5) *Sample Estimates.* To compute these two types of discriminant scores, it is necessary to know the prior probabilities and the parameters of all K density functions. In practice, these quantities will not be known and estimates from initial samples whose classifications are known must be used. It is assumed that the density functions are multivariate normal and that the population parameters may be replaced by unbiased statistics. Thus, in the case of unequal covariance matrices, an "estimated quadratic discriminant score" for the i^{th} population would be defined as

$$\hat{Q}_i(x) = -1/2[\ln \mid \hat{\Sigma}_i \mid] - 1/2(x-\hat{\mu}_i)'\hat{\Sigma}_i^{-1}(x-\hat{\mu}_i) + \ln(\hat{\pi}_i)$$

and used in an approximate Bayes procedure. In a similar fashion, if the covariance matrices are considered equal, an "estimated equivalent linear discriminant score,"

$$\hat{L}_i(x) = -1/2[\ln \mid \hat{\Sigma} \mid] - 1/2(x-\hat{\mu}_i)'\hat{\Sigma}^{-1}(x-\hat{\mu}_i) + \ln(\hat{\pi}_i) \quad i = 1, \cdots, K,$$

is used for classification. These equivalent linear discriminant scores, rather than the linear discriminant scores, are used by the program since the computations for $\hat{Q}_i(x)$ and $\hat{L}_i(x)$ are similar.

There are four major parts to this program: (1) estimation of the density functions parameters from initial samples; (2) testing the assumption of equal covariance matrices in the K populations; (3) computation of discriminant scores and classification of individuals in follow-up samples; and (4) the summarization of the results for both quadratic and linear cases by means of classification matrices. Initial samples need not be the same as those (follow-up) samples classified. The user can specify prior probabilities if they differ from the observed relative frequencies in K initial samples.

Output. The program provides the following output: (a) Within covariance matrix for each initial sample; (b) latent roots of all within covariance matrices; (c) pooled (common) covariance matrix of the initial samples; (d) latent roots of the pooled covariance matrix; (e) mean vector for each initial sample and the overall mean vector for the initial samples; (f) Eigenvectors for all within and pooled covariance matrices; (g) value of the approximate chi-square statistic to test the assumption of homogeneity of covariance matrices; (h) degrees of freedom and associated probability of the chi-square test; (i) sizes and relative frequencies of the K initial samples; (j) quadratic and linear discriminant scores for each individual classified in the follow-up samples (optional); (k) classification matrices for both the quadratic and linear solutions, and these matrices expressed as relative frequencies; (l) chi-square test statistic values, degrees of freedom, and associated probabilities for the classification matrices.

Limitations. The program has the following limitations: (a) the number K of populations (groups) must be less than 10, (b) the number of variables (p) must be less than 9, (c) the product of p and k + 1 must be less than 51, and (d) the number of individuals in each group (either initial or follow-up) must be less than 501.

Computer and Language. The program is written in FORTRAN IV and employs several dimensionless subroutines. Although developed for the IBM 7094, versions of this program have also been adapted for use on the IBM 360, IBM 370, CDC 6500, and CDC 6600.

Availability. A listing and/or copy of the source deck, a manual for input instructions, and sample data may be obtained at no cost from Mark C. Fulcomer, Department of Psychology, New York University, 4 Washington Place, New York, New York 10003. The materials desired and the computer to be used should be specified.

REFERENCES

Anderson, T. W. *An introduction to multivariate statistical analysis.* New York: Wiley, 1958.

Fisher, R. A. The use of multiple measurements in taxonomic problems. Annals of Eugenics, 1936, 7, 179-188.

Fulcomer, M. C. Cross-validation performance of linear and quadratic discriminatory analysis in two populations. Unpublished doctoral dissertation, Ohio State University, 1970.

Rao, C. R. *Linear statistical inference and its applications.* New York: Wiley, 1965.

SICS: Short intradepartmental course scheduler

SALLY A. FLANIK and B. JAMES STARR
Howard University, Washington, D.C. 20001

The SICS program does within-department course scheduling in instructional (e.g., university) settings. Starting with lists of courses that are to be offered and time slots available, the program matches both course and time preferences expressed by the instructors. The likelihood of obtaining desired course and time selections is directly related to the order in which these choices are entered. This system provides maximum flexibility in that the ordering of this input may be varied in accord with the wishes of the faculty. Thus, some departments may wish to give priority to instructors on the basis of rank and length of service, while others may desire matches made either on a first-come, first-served basis or in some other manner.

At the beginning of the process, three course choices and three time selections are solicited from each instructor for each of the *number* of courses for which he is responsible. (To avoid scheduling more than one course at a particular time for a specific instructor, it is important to avoid overlap in the time choice list for any instructor having more than one course scheduled on any single computer run).

The program accepts a parameter indicating the maximum number of classes to be scheduled at any one time. SICS begins processing by reading in the time slots which are available for courses and the courses and sections to be taught for the semester involved. The courses and sections are entered numerically in ascending order of course number (and within courses of section number). Next, the choices of one instructor (for one of his course responsibilities) are processed. Course choices have priority over time choices (i.e., course selections are matched first). When course and time are processed (either by their assignment or via an indication of inability to fill the options), the next instructor's selections are processed.

While some scheduling may still be done by hand, use of SICS has significantly reduced the amount of time that faculty members must spend in course scheduling within this department.

Limitations. Currently, the program operates with a maximum of 100 course-section combinations and 100 time slots. Each instructor may have three choices for course and for time. There is an option for declaring a time slot as "required." This allows only one course (e.g., departmental seminar) to be scheduled at a particular hour. The option is activated on a course card and thus the program selects the time slot.

Computer and Language. Similar programs have been prepared for and run on IBM 1620 and IBM 360-50 computers in FORTRAN IV.

Input. Card input is composed of a number of segments. First, a parameter card provides basic information on days and day sequences for courses. A second parameter card provides information on the maximum number of classes which may be assigned during any one time slot. A subsequent set of cards provides information on the time slots available (to be filled). This is followed by a set of cards indicating courses to be offered. Finally, information in provided regarding the instructors and their choices. Because many of these pieces of information vary little from term to term (for any department), sets of input cards can often be prepared beforehand and revised and reused as needed.

Output. Output consists of a schedule showing course, section, days, time, and instructor. This is followed by a list of instructors without courses (if any exist), and a similar list of courses without instructors. SICS then produces a list of time slots still available for use.

Availability. A program listing and documentation are available from Sally A. Flanik, Department of Psychology, Howard University, Washington, D.C. 20001. The program is available at a cost of $2 (to cover postage and handling) for either a listing or a tape of the program. The tape used may be standard or nonstandard label, 9-track, 800 BPI. The tape will be written with BCD code and should be supplied by the purchaser. Please make checks payable to Sally A. Flanik.