Management Services: A Magazine of Planning, Systems, and Controls

Volume 1 | Number 1 Article 8

3-1964

Discriminant Analysis

Sydney I. Neuwirth

Michael Shegda

DISCRIMINANT ANALYSIS

Presenting a new statistical technique which makes it possible to classify individual members of a group—such as credit applicants—and assign them to a clearly defined part of the whole

by Sidney I. Neuwirth and Michael Shegda
Lybrand, Ross Bros. & Montgomery

CLASSIFICATION of items into distinct groups is an acute problem for management in many companies. For optimal results, the classification should not be based on random choice but rather, upon a systematic methodology.

Discriminant analysis is a technique by which individuals can be classified into categories on a systematic basis. Basically, it provides a statistical means of separating individuals into two or more groups based upon an analysis of their characteristics.

Applications

There are many different areas in which discriminant analysis can be applied—almost any decision-making process which involves classification into two groups as a basis for the decision. Below are three different applications, among the many possibilities:

Credit—Applicants for credit at a bank, finance company, department store, etc., can be classified into "good risks" and "bad risks." Such factors as income, job classification, time at present address, etc., are analyzed for the historically "good" and "bad" applicants to develop a profile of the respective groups. The profile of a new credit applicant is compared to the profiles of good and bad applicants and an assignment is made, i.e., accept or reject the applicant.

Guidance—Many tests of achievement and performance are performed on students in college or the new employee in industry. The scores on these tests and numerical measures of achievement can be correlated with success or failure. Discriminant analysis provides a means for selecting potentially successful persons and, when used on a periodic basis, may provide early warning about students or employees who are not likely to make the grade.

Casualty Insurance—The casualty insurance company which writes automobile insurance is often faced with the problem of classification of driver-applicants in view of differences in premium for drivers with differing driving experiences and backgrounds. Under a "merit-rating" system, automobile bodily injury (B. I.) and property damage (P. D.) premiums can range from 25 per cent below standard for the preferred risk to 25 per cent above standard for the "accident prone" or "substandard" risk. Hence, the driver-applicant has to be classified in many cases prior to the determination of automobile B. I. and P.D. premiums. Based on key characteristics of the driver-applicant's history, a scoring system, developed through the use of discriminant analysis, may be developed which will permit rapid and accurate classification.

As indicated, discriminant analysis is a statistical technique which provides a mechanism by which a population can be separated into two parts.

Each member of a population is defined by a set of characteristics. For example, in a credit application the set of characteristics may include such things as "Age," "Time at Present Address," "Time at Present Occupation," "Monthly Income," and others. The composite of these characteristics represents a profile of the applicant. If the population is composed of two distinct parts, specifically, if the population of credit applicants can be subdivided into two parts, good risks and bad risks, discriminant analysis provides us with a mathematical way to determine the relative weight or importance of each characteristic so that the resultant profile of each of the two classes are at opposite ends of a scale.

X		X
Bad Profile	Scale	Good Profile

Figure 1

To provide a usable methodology, these profiles are translated into numerical scores. The basic equation for determining a score requires only multiplication and addition and is of the form shown.

Score = (Weight of Characteristic 1) x (Value of Characteristic 1)
+ (Weight of Characteristic 2) x (Value of Characteristic 2)
+ (Weight of Characteristic 3) x (Value of Characteristic 3)
+ (Weight of Characteristic 4) x (Value of Characteristic 4)
+
.
.
.
.
+ (Weight of Characteristic n) x (Value of Characteristic n)

The above computation may be best illustrated by an example from the credit field. A balanced sample of historically good risks (paid-up loans) and historically bad risks (charged-off loans) are selected. The characteristics measured are:

Characteristic 1: Age (years)
Characteristic 2: Time at Present Address (months)
Characteristic 3: Time at Present Job (months)
Characteristic 3: Income (Per month)

As a result of the discriminant analysis, the following "weights" are determined to effect maximum separation of the two groups:

Characteristic	*Weight*
1—Age	0.1
2—Time at Present Address	8.2
3—Time at Present Job	7.3
4—Income	2.0

Therefore, the final equation for determining a score is:

Score = (0.1) x (Value of Characteristic 1)
+ (8.2) x (Value of Characteristic 2)
+ (7.3) x (Value of Characteristic 3)
+ (2.0) x (Value of Characteristic 4)

Characteristic	*Average*	
	Good Risk	*Bad Risk*
Age	38	34
Time at Present Address	88	40
Time at Present Job	97	48
Income	416	339

The average values of each of the characteristics are as shown above. The score for the average good risk is:

$$\begin{aligned}\text{Score} &= (0.1) \times (38) + (8.2) \times (88) + (7.3) \times (97) + (2.0) \times (416)\\ &= 3.8 + 721.6 + 708.1 + 832.0\\ &= 2265.5\end{aligned}$$

The score for the average bad risk is:

$$\begin{aligned}\text{Score} &= (0.1) \times (34) + (8.2) \times (40) + (7.3) \times (48) + (2.0) \times (339)\\ &= 3.4 + 328.0 + 350.4 + 678.0\\ &= 1359.8\end{aligned}$$

Hence, we have a score scale similar to Figure 1, except that it is in quantitative terms:

X		X
1359.8 Average Score of Bad Risk	Score Scale	2265.5 Average Score of Good Risk

Figure 2

Since the weights and average scores are based on data from a sample of historically good and bad risks, there will be variation around these average scores. Such variation is normal and provides us with a means for verifying the significance of the average scores.

Since there is some overlap in the distributions, e.g., shaded area in Figure 3, it is necessary to test whether the difference between the average scores for the bad and good risks can be attributed to chance alone or whether this difference is real. If the latter is true, then we have established the technical soundness of the "weighting" and the score scale. This verification is

Computers put this tool within easy reach . . .

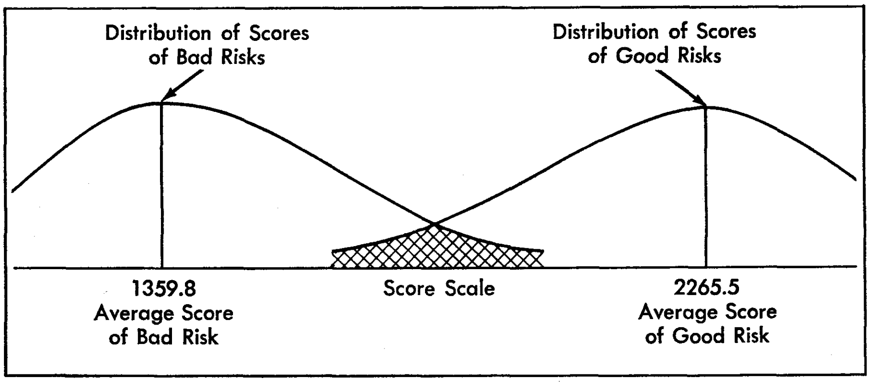

Figure 3

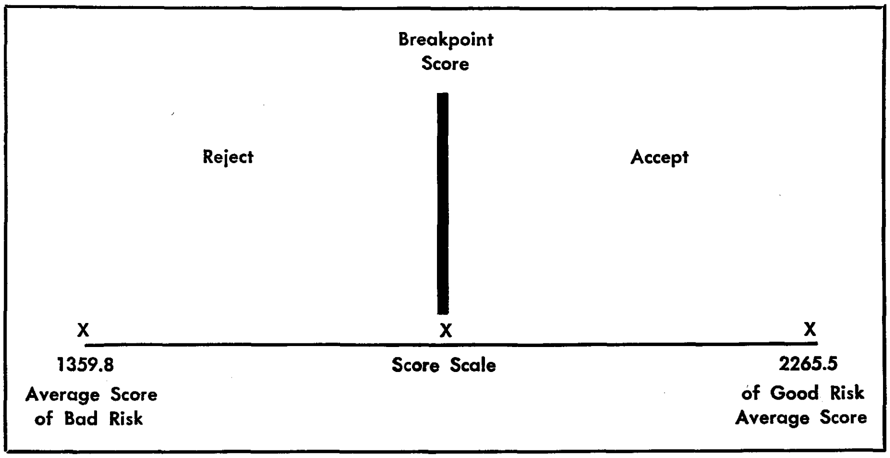

Figure 4

accomplished by application of statistical theory, namely, a statistical test of significance.

What is "breakpoint" score?

Once having established the technical soundness of discriminant analyis applied to a particular problem, it is important to determine how the score will be used in classifying a new member entering the system. In the case of credit applicants, the more specific question is "What is the 'breakpoint' score, above which an applicant is accepted and below which an applicant is rejected?" This is illustrated in Figure 4:

This breakpoint is not necessarily midway between the two averages. The key factors in determining the breakpoint are:

a. Potential gain for correctly classifying a member of the population that had been originally misclassified.

b. Potential loss for incorrectly classifying a member of the population that had been classified correctly originally.

If the potential loss for a misclassification is equal to the potential gain for a correct classification, the breakpoint would be midway between the two averages. However, every application of discriminant analysis must be judged in terms of its own characteristics.

To illustrate further the breakpoint analysis, let us consider credit applicants again. If we examine known good risks (historically) and known bad risks (historically), the key factors in the breakpoint analysis would be:

a. Potential gain for correctly classifying a credit applicant that actually defaulted is the dollar value of the loan or the outstanding principal at the time of default plus any follow-up costs.

b. Potential loss for incorrectly classifying a credit applicant that actually repaid is the interest on the loan.

In this case, it can be seen that there is no equivalence between the potential loss and potential gain; on a dollar basis, there is a greater loss by misclassifying a bad applicant than by misclassifying a good applicant. Hence, the breakpoint is not midway between the average scores for good and bad applicants.

Simulation techniques used

Breakpoint analysis is performed by means of simulation techniques. The score for each of the historically good and bad applicants is determined. Various breakpoint scores are tested in a logical sequence and the potential dollar gains and losses are determined for each. The most favorable breakpoint is selected.

Performing a breakpoint analysis permits us to establish the working mechanism of the system as well as assess the potential dollar benefits. The latter bears on the question of economic feasibility. The breakpoint selected must provide for significant improvements over current company experience.

Discriminant analysis has emerged as a powerful new management decision-making tool. Various applications have been mentioned in this report; however, this technique should be considered applicable wherever the question of classification is the key to a correct management decision. The existence of powerful computer programs for the discriminant analysis and other evaluations put this tool within easy reach of a potential user.

STEPS IN A FEASIBILITY STUDY

I. Feasibility Study

1. From historical records, random samples of equal size from each of the two classifications of interest are extracted, e.g., in credit, this would be equal random samples from known repaid (good risks) and known charged-off (bad risk) applicants.

2. Statistical tests to determine the characteristics which contribute to mathematical separation of the two classes are conducted.

3. Discriminant analysis solution with the use of a computer is performed.

a. "Weights" for characteristics are calculated.

b. Average scores are determined.

c. Statistical tests of technical feasibility are performed.

4. Statistics are collected on potential gains for correct classification and potential losses for incorrect identification, e.g., in credit this would be charged-off dollars and interest dollars for a particular time period.

5. Breakpoint analysis by simulation is performed on computer.

a. Breakpoints are determined.

b. Advantages over current practices are evaluated.

II. Installation

Once having established the soundness of the system on technical, economic and/or other appropriate grounds, the installation phase requires the same steps initially as the feasibility study except that the sample is usually larger. In addition it will require:

1. Design of forms and plans for integrating scoring system within the framework of company's operating policies.

2. Pilot test of system on limited basis.

3. Evaluation of pilot test and modifications in system, if required.

4. Full-scale implementation.

5. Procedures for updating system based on developing experience.