DISCRIMINANT ANALYSIS OF STUDENT LOAN APPLICATIONS

By Edward A. Dyl and Anthony F. McGann

Financial aid officers at colleges and universities are understandably concerned about the number of students who default on their loans. As John H. Mathis noted in a recent in this *Journal* (1),

> . . . Lendable funds are not inexhaustible. They must be repaid after they have served their first generation if succeeding generations are to reuse them.

Mathis' conclusion was that student loan applications must be more accurately evaluated. Universities have, however, been slow to adopt the sophisticated techniques long employed by finance companies and commercial banks to discriminate between good and bad credit risks.[1] This state of affairs is particularly surprising because almost every university with a computer has the resources at hand to develop a credit scoring model tailored to the characteristics of their particular group of student loan applicants.

The purpose of this article is to explain the use of discriminant analysis in identifying potentially "good" versus potentially "bad" student loans. The application of the technique to a sample of 200 student loan applications at the University of Wyoming is demonstrated, and the results are analyzed. The article concludes with some comments about how the reader can apply this technique to his/her own institution.

Edward A. Dyl and Anthony F. McGann are Associate Professors of Business Administration at The University of Wyoming. Professor Dyl received a B.A. degree from Claremont Men's College and M.B.A. and Ph.D. degrees from Stanford University. Professor McGann received a B.S. degree from the United States Military Academy and M.B.A. and Ph.D. degrees from the University of Missouri.

1 The use of discriminant analysis for credit scoring in financial institutions is described in Myers and Forgy (3), Smith (5), Morris (2), and Weingartner (7). Although both Spencer (6) and Pattillo and Wiant (4) have applied statistical techniques to student loan applications, neither provide the basis for a comprehensive model that identifies good versus bad loans. Numbers in parentheses refer to bibliography.

Analysis and Findings

Multivariate discriminant analysis is a statistical technique for classifying an observation (e.g., a loan application) into one of two or more mutually exclusive categories (e.g., good versus bad) based on the observation's individual characteristics. To apply the technique, data are collected concerning potentially relevant (i.e., discriminating) characteristics of each observation and the discriminant analysis model determines the linear combination of these characteristics that best discriminates between the two categories. The resulting discriminant function has the following standard form:

$$Z = c_1V_1 + c_2V_2 + \ldots\ c^nV^n,$$

where Z is a single value, or discriminant score, that can be used to classify the observation; c1, c2, . . . cn are discriminant coefficients computed by the model; and V1, V2, . . . Vn are the independent variables (i.e., the characteristics observed).

In deriving a discriminant function, one is necessarily limited by the data available. Table 1, which lists the potential discriminator variables employed in this study, is a fairly complete summary of the data provided by the student loan application form currently used at the University of Wyoming. Additional data would, of course, provide additional potential discriminator variables. For example, Table 2 shows certain applicant characteristics that earlier studies by Pattillo and Wiant (4) and Spencer (6), which were reported in this *Journal,* concluded were significantly related to student loan repayments. Data on these characteristics might well yield additional discriminating variables.

After tabulating the data, a discriminant function was derived using a standard computer program from the University's computer center's files.[2] The resulting model and its statistical characteristics are shown in Table 3. Seven of the potential discriminator variables had statistically nonzero coefficients, and these variables form the basis for the model. The model might be written as

$$Z = .557V_1 - .554V_2 + .207V_3 + .237V_4 - .208V_5 - .143V_6 + .136V_7$$

The Z value that separates good from bad accounts is -.193. That is, if a loan application has characteristics such that its Z score is greater than this value, it would be classified as a potentially good account. If its Z score is less than this value, it would be classified as a bad account. As long as future loan applicants behave in the same manner as those used to derive the model, the model can be used to discriminate between good and bad accounts. Of course, periodically the model should be revalidated to make certain that it continues to have predictive value.

Note that while V_1, V_2 V_6, and V_7 are scalar values (i.e., numbers), V_3, V_4, and V_5 are dummy variables. That is, V_3, V_4, V_5 are equal to one if the applicant has the particular characteristic and equal to zero if he or she does not.

[2] We employed an algorithm that minimizes Wilks' Λ, a common procedure in discriminant analysis. This procedure chooses variables for the discriminant function that maximize the overall multivariate F-ratio for the difference in group centroids. Prior probabilities of group membership were adjusted in proportion to differences in the size of the two groups; .37 and .63 for the bad and good loan repayment histories respectively.

TABLE 1
APPLICANT CHARACTERISTICS ANALYZED

Class	Residence
1. Freshman	18. Apartment
2. Sophomore	19. House
3. Junior	20. Dormitory
4. Senior	21. Room
5. Graduate Student	22. Sorority/Fraternity
College	Financial Characteristics
6. Agriculture	23. On Scholarship?
7. Arts and Sciences	24. Total Income
8. Commerce and Industry	25. Total Indebtedness
9. Engineering	26. Total University Loans
10. Education	Other Characteristics
11. Health Sciences	27. Own Automobile?
12. Graduate Student	28. Amount Owed on Automobile
Personal Characteristics	Loan Characteristics
13. Age	29. Amount desired
14. Marital Status	30. Monthly payment
15. Sex	31. Co-signer?
16. Grade Point Average	32. Do Parents Know?
17. Number of Children	33. Do Parents Approve?

TABLE 2
ADDITIONAL POTENTIALLY RELEVANT APPLICANT CHARACTERISTICS

1. Applicant's estimated summer income.
2. Previous loan of some kind.
3. Do parents have checking account?
4. Do parents have savings account?
5. Parents total annual income.
6. Value of parents' assets.
7. Does applicant have telephone?
8. Age of applicant's automobile.

Factors Positively Related to Repayment

In this discriminant analysis, four of the significant discriminators displayed direct, positive relationships with actual loan repayment behavior:

(1) Students wth high grade point averages were more likely to pay than those with low GPA's;

(2) Married students were more likely to pay than unmarried students;

(3) Engineering majors were more likely to pay than other majors; and

(4) Students who chose high monthly payments were more likely to pay than those who chose low monthly payments.

It does not seem surprising that higher grade point averages are associated with a higher probability of loan repayment. It is suspected that GPA, which may be as much a measure of socialization as of "intelligence" *per se,* is also colinear with other personal characteristics associated with honoring — and repaying — a debt.

Married student borrowers also had a higher than average probability of repayment, as shown by the positive discriminant function coefficient. While married students comprise about a quarter of the sample (23.5%), they represent nearly one third (31.8%) of the group who repaid their short term university loan. There are, of course, numerous possible explanations of this finding. For example, married students may be more mature, and therefore more responsible, than unmarried students. Alternatively, income provided by a working spouse might be the explanatory factor.

The student borrower's academic major also seemed to be a useful discriminator of repayment behaviors in the sample. In the group studied, no engineering major ever defaulted on his/her loan. Arranging the borrowers' academic majors in descending order of their probability of repayment resulted in the following sequence: engineering, graduate student, agriculture, health sciences, commerce and industry, education, and arts and sciences. When academic majors are considered in conjunction with the other discriminators included in the function, however, only the engineering major was a significant determinant. Presumably, its significance was at least partially due to the good job market for engineers during the period covered by the sample, a possibility that demonstrates the need to update the model every few years, since certain conditions, such as the job market, do change over time.

At first, it was considered somewhat surprising that the size of the monthly payment was positively related to repayment. Upon reconsideration, however, several plausible reasons were found for this relationship. First, large monthly payments are perceptually important so they are likely to be budgeted. Second, a borrower who undertakes large payment is probably eager to pay off his/her loan quickly (e.g., because of discomfort with a debt). Finally, a borrower who agrees to a large payment, quick payback loan may do so with the anticipation of a substantial change in future income, such as could be obtained from a summer or permanent job.

Factors Negatively Related to Repayment

Three factors were negatively associated with repayment: the total amount of other university loans; residence in an apartment; and the size of the short-term loan being requested.

Although students frequently assert that is cheaper to band together and live in a "private" apartment than, say, in a dormitory, they may be fooling themselves. Perhaps the student fails to calculate all of the costs of apartment living. Thus, this discriminator coefficient may simply reflect an unexpected (or uncalculated) demand on the borrower's resources. It may also reflect the more amorphous "life style" of apartment dwellers and this may be unfavorably related to short-term loan repayment.

The magnitude of prior indebtedness to the University and the size of the current loan request are unfavorably associated with repayments. Both are measures of the extent to which the student borrower has agreed to bind future income. While it is not argued that loans to pay for college education are imprudent or harmful to the student, it seems that when other factors are controlled, the student borrower who becomes heavily indebted to the University is also less likely to repay these loans than the student borrower whose indebtedness to the University is smaller.

TABLE 3

SUMMARY OF DISCRIMINANT ANALYSIS

Variable (V_i)	Order of Entry	F-ratio to remove	Standardized Discriminant Coefficients (c_i)	$c_i \neq 0$ at $p \leq$
Grade Point Average (V_1)	1	66.23	.557	.001
Amount of Loan (V_2)	2	59.40	—.554	.001
Engineering Major (V_3)	3	9.58	.207	.01
Married (V_4)	4	7.58	.237	.01
Live in Apartment (V_5)	5	9.53	—.208	.01
Total Amount of University Loans (V_6)	6	4.07	—.143	.05
Size of Monthly Payments (V_7)	7	3.62	.136	.06

Overall Discriminant Function Characteristics:
Eigenvalue = 1.065
Canonical Correlation Coefficient = .718
Wilks' Λ = .484 df = 7
X^2 = 141.1 $p \leq .00$

TABLE 4

PREDICTIVE POWER OF MODEL

Actual Result	Predicted Result Repayment	Default	Total
Default	17	57	74
Repayment	111	15	126
TOTAL	128	72	200

Predictions and the Model

To test the model, it was applied to the same sample of 200 loan applications used to derive the model. The results are summarized in Table 4. The discriminant model correctly classified 84 per cent of the loan applications (i.e., 111 that repaid as agreed and 57 that defaulted out of the 200 applications). In other words, if the model had been used to make the loan decisions, 128 of the loans would have been granted and only 17 of the recipients would have defaulted. In fact, all 200 loans were actually approved by the financial aid office at the University of Wyoming, and 74 recipients defaulted. Thus, while the model had bad debts equal to 13.3 per cent of the loans it granted, the financial aid office had bad debts equal to 37 per cent of the loans granted for this particular sample of 200 loans.

A financial aid officer would probably not, however, employ the model as arbitrarily as done in the test. Presumably, he/she would establish a Z score somewhat higher than -.193 for automatic acceptance of the application and a Z score somewhat lower than -.193 for automatic rejection of the application. Applications with Z scores close to -.193 would be considered marginal and would receive more careful scrutiny. Presumably a good financial aid officer would improve on the model's performance by rejecting some of the 17 bad loans that the model accepted and by accepting some of the 15 good loans that the model rejected. There was, of course, no provision for such "judgment calls" in the test.

Conclusion

This article has explained the application of multivariate discriminant analysis to the problem of identifying good versus bad student loans from data available in the loan application. An example based on student loan experience at the University of Wyoming demonstrated the usefulness of the technique. Although each university will presumably require its own unique discriminant function, the development of such a function is a relatively simple matter. At most universities, both computer programs for discriminant analysis and individuals who are experts in the use of these programs (i.e., business professors or statistics professors) are readily available.

REFERENCES

1 John H. Mathis, "Defaults: Lowering Cloud Over the Guaranteed Loan Program." *The Journal of Student Financial Aid,* 3 (March, 1973).

2 Robert A. Morris, "Credit Analysis: An O.R. Approach." *Management Services,* (March-April, 1966).

3 James H. Myers and Edward W. Forgy, "The Development of Numerical Credit Evaluation Systems." *Journal of the American Statistical Association,* 58 (September, 1963).

4 L. Baker Pattillo, Jr., and Harry V. Wiant, Jr., "Which Students Do Not Repay College Loans?" *The Journal of Student Financial Aid,* 2 (May, 1972).

5 Paul F. Smith, "Measuring Risk on Installment Credit," *Management Science,* 11 (November, 1964).

6 Lee E. Spencer, "Risk Measurement for Short Term Loans." *The Journal of Student Financial Aid,* 4 (November, 1974).

7 H. Martin Weingartner, "Concepts and Utilization of Credit-Scoring Techniques." *Banking,* (February, 1966).